

# Psychological Bulletin

---

## CONTENTS

### ARTICLES:

- Content-Analysis Studies of Psychotherapy.....FRANK AULD, JR.  
AND EDWARD J. MURRAY 377
- Employee Attitudes and Employee  
Performance.....ARTHUR H. BRAYFIELD  
AND WALTER H. CROCKETT 396
- The "Post-Mortem" Testing of Experimental  
Comparisons.....RICHARD B. MCHUGH  
AND DOUGLAS S. ELLIS 425
- The Infantile Disorders of Hospitalism and Anaclitic  
Depression.....SAMUEL R. PINNIAU 429
- Reply to Dr. Pinnia.....RENÉ A. SPITZ 453
- Reply to Dr. Spitz.....SAMUEL R. PINNIAU 459

### BOOK REVIEWS:

- Good and Scates's Methods of Research.....MALCOLM G. PRESTON 465
- Brand's The Study of Personality.....BERT R. SAPPENFIELD 468
- Kerman's What Is Electroshock Therapy?.....JAMES D. PAGE 466
- Ryan and Smith's Principles of Industrial Psychology.....MILTON L. BLUM 466
- Osser and Hammond's Social Structure and Personality in a City, and Osser and  
Emery's Social Structure and Personality in a Rural Community...DAN L. ADLER 468
- Burt's The Cause and Treatment of Backwardness.....SEYMOUR B. SARASON 469
- Kugelman's The Management of Mental Deficiency in Children ....KARL F. HEISER 470
- Edwards' Statistical Methods for the Behavioral Sciences.....LEONARD S. KOGAN 470
- Le Beau's Psycho-chirurgie et Fonctions Mentales.....H. J. EYSENCK 471

---

Published Bimonthly by the  
American Psychological Association

WAYNE DENNIS, Editor

Brooklyn College (on leave of absence, 1955-56)

The American University, Washington, D.C. (1955-56)

HOWARD GARDNER, Associate Editor (Book Reviews)

Brooklyn College

ROBERT A. THOMPSON, Associate Editor (Statistics)

Brooklyn College, Graduate Institute

JOSEPHINE BOWEN, Managing Editor

Consulting Editors

LARRY F. CARMICHAEL

Editor, *Journal of Experimental Psychology*

University of California, Berkeley

JOSEPH C. RABY

University of Colorado

BENTON J. UNDERWOOD

Northwestern University

E. RAYNE WALLACE

U.S. Insurance Agency

Psychological Association

*The Psychological Bulletin* contains evaluative reviews of research literature and studies on research methodology in psychology. This Journal does not publish reports of original research or original theoretical articles.

Manuscripts should be sent to Wayne Dennis, Department of Psychology, Brooklyn College, Brooklyn 10, New York. Manuscripts should be sent to this address even during the editor's absence on leave. Manuscripts will be acknowledged at this address and (optional) copies forwarded to the editor. However, beginning in August 1, 1955 and until August 1, 1956 correspondence should be addressed to the editor of the Department of Psychology, American University, Washington, D.C.

Notwithstanding past reviews, for this year, book reviews will no longer appear in the Journal, but in *Contemporary Psychology*, an APA publication starting in January, 1955. Reviews books for review should be sent to E. G. Boring, Memorial Hall, Harvard University, Cambridge 38, Mass.

**Preparation of articles for publication.** Authors are strongly advised to follow the general directions given in the "Publication Manual of the American Psychological Association" (*Psychological Bulletin*, 1952, 58 [Vol. 4, Part 2], 389-449). Special attention should be given to the section on the preparation of the references (pp. 422-429), since this is a particularly common source of difficulty in long reviews of research literature. All references should be placed at the end of the article. All manuscripts should be submitted in duplicate. Original figures should be prepared for publication; duplicate figures may be photographed or made by hand. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

**Reprints.** Fifty free offprints are given to authors of articles, notes, and special reviews. Five copies of the Journal are mailed gratis to the editors of both reviews.

**Communications**—including subscriptions, orders of back issues, and changes of address—should be addressed to the American Psychological Association, 1200 Massachusetts Avenue, N.W., Washington 6, D.C. Address changes must reach the Executive Office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the publisher if they will guarantee payment of forwarding postage. Other claims for undelivered copies must be made within 6 months of publication.

Journal subscription: \$2.00 (Foreign \$2.50). Single copies, \$1.50.

PUBLISHED MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Executive Office

1200 Massachusetts Avenue, N.W., Washington 6, D.C.

Reprints should also be ordered at the same office at Washington, D.C., under the call of Reprints, Inc., Attention: Reprints, 1200 Massachusetts Avenue, N.W., Washington 6, D.C. Reprints are available on request at special rates of purchase for the purchase of 100, 500, or 1000 reprints; for 1000 reprints, the rate is \$1.00 per copy. Reprints are available in 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 2679, 2680, 2681, 2682, 2683, 2684, 2685, 2686, 2687, 2688, 2689, 2690, 2691, 2692, 2693, 2694, 2695, 2696, 2697, 2698, 2699, 2700, 2701, 2702, 2703, 2704, 2705, 2706, 2707, 2708, 2709, 2710, 2711, 2712, 2713, 2714, 2715, 2716, 2717, 2718, 2719, 2720, 2721, 2722, 2723, 2724, 2725, 2726, 2727, 2728, 2729, 2730, 2731, 2732, 2733, 2734, 2735, 2736, 2737, 2738, 2739, 2740, 2741, 2742, 2743, 2744, 2745, 2746, 2747, 2748, 2749, 2750, 2751, 2752, 2753, 2754, 2755, 2756, 2757, 2758, 2759, 2760, 2761, 2762, 2763, 2764, 2765, 2766, 2767, 2768, 2769, 2770, 2771, 2772, 2773, 2774, 2775, 2776, 2777, 2778, 2779, 2780, 2781, 2782, 2783, 2784, 2785, 2786, 2787, 2788, 2789, 2790, 2791, 2792, 2793, 2794, 2795, 2796, 2797, 2798, 2799, 2800, 2801, 2802, 2803, 2804, 2805, 2806, 2807, 2808, 2809, 2810, 2811, 2812, 2813, 2814, 2815, 2816, 2817, 2818, 2819, 2820, 2821, 2822, 2823, 2824, 2825, 2826, 2827, 2828, 2829, 2830, 2831, 2832, 2833, 2834, 2835, 2836, 2837, 2838, 2839, 2840, 2841, 2842, 2843, 2844, 2845, 2846, 2847, 2848, 2849, 2850, 2851, 2852, 2853, 2854, 2855, 2856, 2857, 2858, 2859, 2860, 2861, 2862, 2863, 2864, 2865, 2866, 2867, 2868, 2869, 2870, 2871, 2872, 2873, 2874, 2875, 2876, 2877, 2878, 2879, 2880, 2881, 2882, 2883, 2884, 2885, 2886, 2887, 2888, 2889, 2890, 2891, 2892, 2893, 2894, 2895, 2896, 2897, 2898, 2899, 2900, 2901, 2902, 2903, 2904, 2905, 2906, 2907, 2908, 2909, 2910, 2911, 2912, 2913, 2914, 2915, 2916, 2917, 2918, 2919, 2920, 2921, 2922, 2923, 2924, 2925, 2926, 2927, 2928, 2929, 2930, 2931, 2932, 2933, 2934, 2935, 2936, 2937, 2938, 2939, 2940, 2941, 2942, 2943, 2944, 2945, 2946, 2947, 2948, 2949, 2950, 2951, 2952, 2953, 2954, 2955, 2956, 2957, 2958, 2959, 2960, 2961, 2962, 2963, 2964, 2965, 2966, 2967, 2968, 2969, 2970, 2971, 2972, 2973, 2974, 2975, 2976, 2977, 2978, 2979, 2980, 2981, 2982, 2983, 2984, 2985, 2986, 2987, 2988, 2989, 2990, 2991, 2992, 2993, 2994, 2995, 2996, 2997, 2998, 2999, 3000, 3001, 3002, 3003, 3004, 3005, 3006, 3007, 3008, 3009, 3010, 3011, 3012, 3013, 3014, 3015, 3016, 3017, 3018, 3019, 3020, 3021, 3022, 3023, 3024, 3025, 3026, 3027, 3028, 3029, 3030, 3031, 3032, 3033, 3034, 3035, 3036, 3037, 3038, 3039, 3040, 3041, 3042, 3043, 3044, 3045, 3046, 3047, 3048, 3049, 3050, 3051, 3052, 3053, 3054, 3055, 3056, 3057, 3058, 3059, 3060, 3061, 3062, 3063, 3064, 3065, 3066, 3067, 3068, 3069, 3070, 3071, 3072, 3073, 3074, 3075, 3076, 3077, 3078, 3079, 3080, 3081, 3082, 3083, 3084, 3085, 3086, 3087, 3088, 3089, 3090, 3091, 3092, 3093, 3094, 3095, 3096, 3097, 3098, 3099, 3100, 3101, 3102, 3103, 3104, 3105, 3106, 3107, 3108, 3109, 3110, 3111, 3112, 3113, 3114, 3115, 3116, 3117, 3118, 3119, 3120, 3121, 3122, 3123, 3124, 3125, 3126, 3127, 3128, 3129, 3130, 3131, 3132, 3133, 3134, 3135, 3136, 3137, 3138, 3139, 3140, 3141, 3142, 3143, 3144, 3145, 3146, 3147, 3148, 3149, 3150, 3151, 3152, 3153, 3154, 3155, 3156, 3157, 3158, 3159, 3160, 3161, 3162, 3163, 3164, 3165, 3166, 3167, 3168, 3169, 3170, 3171, 3172, 3173, 3174, 3175, 3176, 3177, 3178, 3179, 3180, 3181, 3182, 3183, 3184, 3185, 3186, 3187, 3188, 3189, 3190, 3191, 3192, 3193, 3194, 3195, 3196, 3197, 3198, 3199, 3200, 3201, 3202, 3203, 3204, 3205, 3206, 3207, 3208, 3209, 3210, 3211, 3212, 3213, 3214, 3215, 3216, 3217, 3218, 3219, 3220, 3221, 3222, 3223, 3224, 3225, 3226, 3227, 3228, 3229, 3230, 3231, 3232, 3233, 3234, 3235, 3236, 3237, 3238, 3239, 3240, 3241, 3242, 3243, 3244, 3245, 3246, 3247, 3248, 3249, 3250, 3251, 3252, 3253, 3254, 3255, 3256, 3257, 3258, 3259, 3260, 3261, 3262, 3263, 3264, 3265, 3266, 3267, 3268, 3269, 3270, 3271, 3272, 3273, 3274, 3275, 3276, 3277, 3278, 3279, 3280, 3281, 3282, 3283, 3284, 3285, 3286, 3287, 3288, 3289, 3290, 3291, 3292, 3293, 3294, 3295, 3296, 3297, 3298, 3299, 3300, 3301, 3302, 3303, 3304, 3305, 3306, 3307, 3308, 3309, 3310, 3311, 3312, 3313, 3314, 3315, 3316, 3317, 3318, 3319, 3320, 3321, 3322, 3323, 3324, 3325, 3326, 3327, 3328, 3329, 3330, 3331, 3332, 3333, 3334, 3335, 3336, 3337, 3338, 3339, 3340, 3341, 3342, 3343, 3344, 3345, 3346, 3347, 3348, 3349, 3350, 3351, 3352, 3353, 3354, 3355, 3356, 3357, 3358, 3359, 3360, 3361, 3362, 3363, 3364, 3365, 3366, 3367, 3368, 3369, 3370, 3371, 3372, 3373, 3374, 3375, 3376, 3377, 3378, 3379, 3380, 3381, 3382, 3383, 3384, 3385, 3386, 3387, 3388, 3389, 3390, 3391, 3392, 3393, 3394, 3395, 3396, 3397, 3398, 3399, 3400, 3401, 3402, 3403, 3404, 3405, 3406, 3407, 3408, 3409, 3410, 3411, 3412, 3413, 3414, 3415, 3416, 3417, 3418, 3419, 3420, 3421, 3422, 3423, 3424, 3425, 3426, 3427, 3428, 3429, 3430, 3431, 3432, 3433, 3434, 3435, 3436, 3437, 3438, 3439, 3440, 3441, 3442, 3443, 3444, 3445, 3446, 3447, 3448, 3449, 3450, 3451, 3452, 3453, 3454, 3455, 3456, 3457, 3458, 3459, 3460, 3461, 3462, 3463, 3464, 3465, 3466, 3467, 3468, 3469, 3470, 3471, 3472, 3473, 3474, 3475, 3476, 3477, 3478, 3479, 3480, 3481, 3482, 3483, 3484, 3485, 3486, 3487, 3488, 3489, 3490, 3491, 3492, 3493, 3494, 3495, 3496, 3497, 3498, 3499, 3500, 3501, 3502, 3503, 3504, 3505, 3506, 3507, 3508, 3509, 3510, 3511, 3512, 3513, 3514, 3515, 3516, 3517, 3518, 3519, 3520, 3521, 3522, 3523, 3524, 3525, 3526, 3527, 3528, 3529, 3530, 3531, 3532, 3533, 3534, 3535, 3536, 3537, 3538, 3539, 3540, 3541, 3542, 3543, 3544, 3545, 3546, 3547, 3548, 3549, 3550, 3551, 3552, 3553, 3554, 3555, 3556, 3557, 3558, 3559, 3560, 3561, 3562, 3563, 3564, 3565, 3566, 3567, 3568, 3569, 3570, 3571, 3572, 3573, 3574, 3575, 3576, 3577, 3578, 3579, 3580, 3581, 3582, 3583, 3584, 3585, 3586, 3587, 3588, 3589, 3590, 3591, 3592, 3593, 3594, 3595, 3596, 3597, 3598, 3599, 3600, 3601, 3602, 3603, 3604, 3605, 3606, 3607, 3608, 3609, 3610, 3611, 3612, 3613, 3614, 3615, 3616, 3617, 3618, 3619, 3620, 3621, 3622, 3623, 3624, 3625, 3626, 3627, 3628, 3629, 3630, 3631, 3632, 3633, 3634, 3635, 3636, 3637, 3638, 3639, 3640, 3641, 3642, 3643, 3644, 3645, 3646, 3647, 3648, 3649, 3650, 3651, 3652, 3653, 3654, 3655, 3656, 3657, 3658, 3659, 3660, 3661, 3662, 3663, 3664, 3665, 3666, 3667, 3668, 3669, 3670, 3671, 3672, 3673, 3674, 3675, 3676, 3677, 3678, 3679, 3680, 3681, 3682, 3683, 3684, 3685, 3686, 3687, 3688, 3689, 3690, 3691, 3692, 3693, 3694, 3695, 3696, 3697, 3698, 3699, 3700, 3701, 3702, 3703, 3704, 3705, 3706, 3707, 3708, 3709, 37

# Psychological Bulletin

## CONTENT-ANALYSIS STUDIES OF PSYCHOTHERAPY

FRANK AULD, JR. AND EDWARD J. MURRAY<sup>1</sup>

*Yale University*

It is only sixty years since Freud developed psychoanalysis (40, p. 252) and thereby originated dynamic psychotherapy. Although psychoanalysts and other investigators have learned a great deal about psychotherapy in these sixty years, studies of psychotherapy have suffered from three hindrances:

1. The basic data of psychotherapy were transient. Furthermore, they were accessible only to the therapist; others had to take his word for what happened in the interviews. The consequences of this lack of adequate recording are eloquently set forth by Kubie (44).

2. Conclusions stemming from investigations were matters of impression and opinion, because there was no technique for studying the verbal materials objectively.

3. The data could not be fitted into a suitable theoretical framework because the psychologies of the day (for example, Wundt's and Brentano's) had little relevance to the phenomena observed. To appreciate this, one might imagine oneself in

Freud's place in 1895 and ask, "How could I explain the things I've observed?"

Recent methodological and theoretical developments seem to justify the hope that these three obstacles to scientific research on psychotherapy can be overcome. The new methods are sound recording and content analysis; the new theoretical developments are the recent attempts to develop a general science of behavior.

*Sound recording* of interviews has made a common set of data available to scientists, a set that can be preserved and be studied as many times as necessary (7, 26, 76). Pioneers in the recording of psychotherapy interviews include Zinn, Lasswell (45, 46, 47), Rogers (80, 81), Robinson (79), Covner (14, 15, 16), and Porter (70, 71). All of the studies reviewed in this article, with one exception, derived their data from sound recordings.

*Content analysis* is a method for studying the content of communication in an objective, systematic, and quantitative way (Berelson, 4, p. 18). "Content" is what is said. Lasswell (48, 49) pioneered the application of content analysis to social science problems. In the past decade content analysis has been widely used for studying therapeutic interviews. All the studies reviewed in this paper are content-analysis studies.

<sup>1</sup> The authors wish to thank Professor John Dollard for his guidance and support of their research. Professor Dollard read this paper and made helpful suggestions. This paper is a product of research project M-648, "Development of quantitative methods for detailed study of psychotherapy," supported by the National Institute of Mental Health, U. S. Public Health Service.

*General theories of behavior*, which have been developed in recent years, provide possible frameworks for explaining the data of psychotherapy and for guiding investigators toward studies that will advance our understanding of human behavior as well as give answers to specific questions (32). The fusion of psychological theory with psychotherapeutic research is exemplified in the work of Dollard and Miller (20). Although the non-directive group at first had no comprehensive theoretical platform, in recent years these investigators have shown a growing tendency to state explicit theories and to design their investigations to test hypotheses derived from the theories (83, 84). And, of course, it cannot be forgotten that the psychoanalysts, building on their clinical experience, constructed a comprehensive psychological theory. Despite its vagueness and lack of rigorous formulation, it has served as a guide for a number of investigators.

In this paper we attempt to review the considerable body of literature on content analysis of recorded interviews.

#### A SURVEY OF THE STUDIES

The content-analysis studies of psychotherapy fall into three general classes: (a) Methodological studies, in which the aim was principally to develop measures, (b) Descriptive studies of cases, and (c) Theoretically guided studies of therapy (i.e., studies of cause-and-effect relationships). Although it is admitted that the third kind of study, if well done, contributes the most to our understanding, studies of the first type are necessary to develop methods and studies of the second type may provide hunches that can be rigorously tested later by a theoretically oriented study.

#### *Methodological Studies*

In these studies, the emphasis is on development of new measures rather than on testing of hypotheses. Nevertheless, the inventor of a measure always has some theoretical presuppositions; and these are what influence him to measure one kind of thing rather than another. Porter (70, 71), for instance, developed a classification of therapist responses. His classification emphasizes the degree to which the therapist takes responsibility for the interview, which Porter considers an important variable in therapy (72, pp. 45-60). Among the nondirective therapists, Royer (87), Snyder (91), Curran (17), Raimy (73), Hogan (34), Haigh (31), Stock (93), Hoffman (33), Kahn (41), and others have developed measures of various kinds. We shall choose several measures for special comment, because they illustrate the problems involved in such classifications.

*Snyder's system.* Snyder's system (91) has been especially influential. Seeman (88), Aronson (1), Rakusin (74), Tucker (96), Gillespie (27), and Blau (6) have made use of it in their investigations. In Snyder's system, the therapist's responses are classified according to a modification of Porter's categories. The categories designate the technique which the therapist uses: restating content, clarifying feeling, interpreting, structuring, leading, suggesting, questioning, persuading, accepting, reassuring, approving, disapproving. The client's responses are classified under these headings: problems, simple responses (questions, answers, acceptance, disagreement), insight, planning. Notice that Snyder, like Porter, considers the degree of responsibility assumed by the therapist to be an important variable in the therapist's behavior. On the side of the client, Snyder's



classification implies that it is important to notice whether the client is discussing problems or showing understanding. Presumably in successful cases problems dominate at the beginning and insight and plans dominate at the end.

*Curran's two systems.* Curran (17) has reported a detailed analysis of a single case. His study is of special interest for two reasons. First, he measured "insight" by noting instances of the client's connecting two different problems. Also of interest is Curran's classification of the problems discussed by the client. The problems discussed by the client included: hostility, dependency, insecurity, unhappiness, conflict, discouragement, withdrawal, daydreaming, feelings of inferiority, sex, sin, younger brother, school work, and war. The special interest of this classification lies in the fact that while most researchers have only noted whether the client talked about a problem or about something else, Curran described *what* the problems were that the client talked about.

*Haigh's measure of defensiveness.* Haigh (31) developed a measure of defensiveness, building on previous work by Hogan (34). Haigh defined defensive behavior as behavior that

shifts attention away from an inconsistency that, if perceived, would threaten the client.<sup>3</sup> It may be noted, by the way, that in order to judge a remark as defensive according to Haigh's system one must make the judgment that two of the statements of the client are incongruent or that the client "ought" to be exploring some aspect of his emotional life and is avoiding it (31, p. 182).

*Interaction Process Analysis.* The theory behind this system, which was developed by Bales (3), is that problem solving involves a series of steps: getting information, making decisions, carrying out actions. At each stage of problem solving, the participants interact with each other. One participant, for instance, asks for information; another gives information; then a third member of the group may offer an opinion of the correctness of this information; the individual who gave the information may then proceed to defend his views with some heat. As applied to therapy, the Bales system is used to score both the remarks of the therapist and those of the client.

*Discomfort-Relief Quotient.* The Discomfort-Relief Quotient, or D.R.Q., was suggested by behavior theory. Behavior theory states that responses are incited by drives and reinforced by drive reduction. The learning of new habits is, according to this theory, accompanied by drive reduction. The new learning that occurs in successful therapy ought, therefore, to be accompanied by a reduction in drive. Guided by these theoretical notions, Dollard and Mowrer (21) attempted to measure the amount of drive borne by the client. They did this by classifying each word as a drive, reward, or neutral word according as the word seemed to represent discomfort (suf-

<sup>3</sup> This definition seems essentially the same as the psychoanalytic definition of "resistance." Colby, for example, defines resistances as "those defenses that operate in and against the therapeutic process to prevent an uncovering and a dissolution of the neurotic conflict" (12, p. 8). Haigh's definition differs from the analytic one, however, in this respect: He notes only those resistances that can be easily inferred from the content of the client's speech, i.e., from the client's denials, rationalizations, and projections. Analysts, on the other hand, note not only these but also such evidences of resistance as: silences, avoidance of obvious topics, lateness in keeping appointment, and various transference responses (12, pp. 95-106).

fering, tension, pain, unhappiness) or relief (comfort, satisfaction, enjoyment) or neither. The D.R.Q. is obtained by dividing the number of discomfort words by the total number of discomfort and relief words. The authors demonstrated very satisfactory reliability of D.R.Q. scores. Dollard and Mowrer, employing other units besides words, found that sentences and "thought units" could also be reliably scored.

Dollard and Mowrer intended the D.R.Q. as a measure of the tension *experienced* by the client. We know, however, that the client's verbal responses do not always accurately label his drives (20). It is possible that: (a) A verbal response may indicate a drive when the drive is not present, as when a client makes an insincere complaint in order to enlist the sympathy of the therapist. (b) A drive may be present with no verbal response describing it, as when a client denies having sexual wishes while his other behavior convinces us that he does have them. (c) A verbal response may describe a conflict different from the one the patient is actually experiencing, as when talk about vocational problems replaces talk about an unconscious homosexual conflict. It is also possible that the importance of a drive is not accurately reflected by the *frequency* of occurrence of sentences about it. It is the task of empirical research to discover whether clients accurately label their drives or whether these possible distorting factors interfere with accurate labeling to such a degree that verbal responses cannot be used as indices of drive.

In a study of interviews from six recorded psychotherapy cases, Mowrer *et al.* (62) found a moderately high relationship between the D.R.Q. scores for interviews and measures of palmar sweating made after the same

interviews. For five of these cases, the  $r$  between D.R.Q. and palmar sweating was .57; in the sixth case, for which the correlation was computed separately, the  $r$  was only .30. To the extent that palmar sweating is itself an adequate measure of "tension," this finding somewhat strengthens the case for considering the D.R.Q. a measure of tension. Meadow *et al.* (53), studying psychotic patients, found no significant relationship between D.R.Q. score in a special interview and a psychiatrist's rating of tension made on the basis of other interviews with the patient. In interpreting this result one should note, first, that the rating of tension was not made in the same situation in which the D.R.Q. score was obtained; second, that the result of a study of psychotic patients, who as a group are likely to make verbal responses without showing "appropriate" emotional reactions, may not predict very well what will be found when normal and neurotic subjects are studied; and, finally, that the validity of the psychiatrist's ratings is not known.

The problem of how the verbal behavior of the client, which is taken note of in content analysis, is related to his nonverbal behavior, is a difficulty not only for the D.R.Q. but also for every other system of content analysis.

*Positive-Negative-Ambivalent Quotient.* A measure independently developed by Raimy (73) has a strong family resemblance to the D.R.Q. Raimy's PNAvQ is a quotient, as the D.R.Q. is; and, like the D.R.Q., it is an indication of positive and negative emotional reactions of the client. It differs from the D.R.Q. in focusing attention on the statements of the client about himself (whereas the D.R.Q. includes all statements of the client, whether they are self-refer-

ential or not). In some cases, the two measures would be expected to give similar results, which is what Kauffman and Raimy (43) found to be true in a study of 17 counseling interviews. These authors also found the PNAvQ to be somewhat lower, on the average, than the D.R.Q.

Horwitz (35), however, found only a low correlation ( $r=.38$ ) between the D.R.Q. and PNAvQ in 36 initial casework interviews. It is possible that the low correlation should be attributed to unreliability of scoring; but this seems unlikely, since investigators using these measures in other studies have found them to be reliable. It is more likely that the correlation is low because the D.R.Q. includes all statements of the client, whether about himself or about others, while the PNAvQ includes only statements referring to the client. If the client talks a great deal about other persons, these statements will affect the D.R.Q. but not the PNAvQ—with a resulting drop in agreement between the two measures.

Doubting that a measure of the "self concept" would be much related to the D.R.Q. if the "self concept" were narrowly defined, Horwitz also investigated the correlation between the D.R.Q. and a "self-approval quotient." The self-approval quotient is the same as the PNAvQ except that statements which indicate that the client is happy, glad, or improving, or fearful, anxious, or unsuccessful are not included. (Such statements are included in the PNAvQ.) The only statements included in the self-approval quotient are those in which the client evaluates himself. The D.R.Q. was not significantly correlated with the self-approval quotient ( $r=.14$ ).

*A system in terms of motivation and conflict.* E. J. Murray (63, 64), in devising a system for the study of motivation and conflict in psycho-

therapy, was influenced by psychoanalysis and learning theory. Murray wished to designate underlying motivations of the client; but, at the same time, he wanted the system to be objective, i.e., to have explicit rules for inferring the underlying motives. In keeping with the wish to have an objective system, Murray chose categories that require relatively little inference by the scorer and reflect chiefly manifest content. The category system designates statements expressing a need, statements expressing anxiety about a need, and statements expressing hostility on account of frustration of a need. The drives included are: sex, affection, dependence, independence, and "unspecified" drive. The system also requires the scorer to designate the person who is the object of the need, e.g., the person who is loved or hated.

Murray (63, 64) also devised categories for describing the therapist's behavior, building on the theoretical work of Dollard and Miller (20). Remarks of the therapist were classified according to activity-passivity and according to their function as rewards, punishments, labels, discriminations, generalizations, instructions concerning free association, directions, or probes.

The reliability of scoring these categories was studied (63, 64) and found to be fairly high. It was found that reliability decreased as a greater number of simultaneous discriminations were required of the scorer. This finding agrees with the experience of Kaplan and Goldsen (42).

*Lasswell's general purpose system.* Lasswell (48) has suggested a scheme for classifying interview materials under very general headings, so that the system can be applied to interviews differing widely in topic. The classification tells who makes the statement (talker-listener-another),

whom the statement refers to (talker-listener-another), and what attitude is expressed (favorable-unfavorable). To our knowledge this scheme has not been applied to psychotherapy cases, except apparently in a recent study by Rosenman (85).

*Other measures.* Other proposed measures include: Collier's (13) scale for the degree of "uncovering" used by the therapist, Helen E. Miller's (54) measure of "acceptance," Raskin's (75) "locus-of-evaluation" rating, Elton's (22) responsibility rating, Carnes and Robinson's (9) talk ratio, the various measures used by Lewis in her pioneering study (50), and White's (97) frustration-satisfaction ratio, a measure similar to the D.R.Q.

### *Descriptive Studies*

*Movement from problems to insight.* Snyder (91) has described how the content of the client's speech changes during the course of nondirective therapy. According to Snyder, first there is a statement of a problem or problems; then discussion of these problems, with increased insight; and finally, formulation of plans for new responses. Studying four "successful" cases and one "unsuccessful" case, Snyder (92) found that the successful cases showed a trend from problems to plans; the unsuccessful case failed to show this trend. Seeman (88) repeated Snyder's study, using a larger sample of cases, and confirmed his results. Results reported by Snyder (91), by Seeman (88), and by Blau (6) also indicate that in the cases that were believed successful, the client talked less about problems and more about insights and plans at the end of treatment than at the start; similarly, he expressed more positive feelings and fewer negative feelings than he did at the beginning. Such a change was not observed in unsuccessful cases.

This result is hardly an independent confirmation that the cases judged to be successful *were* successful, since the judges undoubtedly reacted to just such changes in the client's verbalizations when deciding whether cases were successful. While these studies do not give independent confirmation, they are nevertheless valuable because they specify some ways in which the client's speech behavior changed, and they provide reliable measures for these changes.

*Attitudes toward self.* Raimy (73) studied the kinds of attitudes toward self expressed by the client at various times in the course of therapy. He found that in cases that were judged to be successful, positive statements about self increase and negative statements decrease during the course of treatment. Seeman (88) studied all attitudes of the client, both about self and about other people, and found that positive expressions increase and negative expressions decrease during therapy. He found no significant correlation between the change from negative to positive attitudes and the counselor's rating of success of the case, if all statements of attitude were considered; but when he examined only statements expressed in the present tense, he found a significant correlation between changes in these statements and the counselor's rating ( $r = .66$ ). Bugental (8) has used Raimy's scales and somewhat modified them.

*Studies using the D.R.Q.* Still another way of describing the course of psychotherapy is to note changes in the degree of tension expressed in the client's sentences. A measure of verbal tension (the D.R.Q.) has already been described in the preceding section, and the theoretical considerations that guided its invention were presented there. Results with the use of the D.R.Q. will be briefly presented here. Hunt (36, 37, 38, 61)

found little agreement between changes in the D.R.Q. (scored from caseworkers' summaries) from the beginning to the end of casework and the degree of movement or progress in the case as judged by experienced caseworkers. When these judgments of experienced caseworkers were quantified and made highly reliable, by use of a scale of "movement," there was still little agreement between movement and drop in D.R.Q. Assum and Levy (2), studying one case, found a drop in D.R.Q. accompanying success. Natalie Rogers (61) studied three cases, one judged as very successful, one as moderately successful, and one as unsuccessful. The successful cases showed a drop in D.R.Q., but the unsuccessful one did not. Cofer and Chance (10) computed the D.R.Q. for each hour of five published cases—four nondirective cases and the case in Lindner's *Rebel Without a Cause* (51). All the cases were judged as successful by the therapists, and all showed a drop in D.R.Q. Mowrer (61) reported one of his own cases which showed a drop in D.R.Q. and was believed to be a successful case.

On the evidence so far, there is no clear relationship between the change in D.R.Q. from beginning to end of a case and the success of the case. Some investigators have reported a drop in D.R.Q. accompanying success (these investigators studied, in total, verbatim transcripts of 10 cases); others (who studied case workers' notes of 38 cases) found no relationship between drop in D.R.Q. and success of the cases. It may be noted that no evidence for the reliability of the judgments of "success" was reported for those 10 cases showing positive results; but reliability of the judgments of "movement" in the 38 cases was established.

Murray, Auld, and White (65) have reported a "partly successful"

case showing no drop in D.R.Q. Their paper includes a discussion of the reasons for absence of any change in the D.R.Q.

The critical issue here, it seems to the present authors, is that we have no adequate measure of "success." Until such a measure has been devised, it makes little sense to ask whether a drop in D.R.Q. is related to success.

*Relations of problems to each other.* Curran (17), studying 20 recorded interviews of a single case, described changes in the client's perceptions of relationships between his problems. At the beginning of therapy, the client discussed a large number of problems and talked about them separately; at the end of therapy he discussed a smaller number of problems, tending to combine problems that had previously been separate. Curran's contribution to methodology is his idea of studying "insight" by noting the client's discovery of relationships between previously separate problems.

*Differences between different therapies.* Porter (70, 71) described differences between "directive" and "nondirective" counselors. The directive counselors studied by Porter talked much more than the nondirective counselors and took more responsibility for guiding the interview by questioning, explaining, probing, and advising. Gump (30) compared responses of the therapists in one psychoanalytic and five nondirective cases and noted differences in technique. No generalizations about analytic or nondirective therapy should be made, however, on the basis of only six cases.

Strupp (94) has compared the verbal responses of Rogerian and non-Rogsonian therapists when they were asked what they would say after various statements by a client. Answers of the therapists were classified ac-



cording to Bales' system, with fairly high reliability (78% agreement between two scorers). As would be expected, there were differences between the answers of Rogerian and non-Rogerian therapists. Variation within the two groups of therapists was also large. Strupp (95) further classified the non-Rogerian therapists according to whether they had received a personal analysis. Those who had been analyzed were more likely to give answers classified as "disagrees" and "shows antagonism" and less likely to give answers coded as "agrees, shows passive acceptance." The analyzed therapists less often indicated that they would say nothing at all after the client's statement.

*Group psychotherapy.* Coffey *et al.* (11) have presented a description of the content of group psychotherapy interviews with college students. The nuclear conflict of students in these groups was described by the authors as follows: "High standards and intellectualized ideals are often associated with passivity and inhibition of emotions. The constricted feelings find distorted expression in intellectualization and isolated fantasy resulting in varying amounts of guilt, immobilization, and affective sterility" (11, p. 58). A classification of what the groups talked about yielded the following results: Sexuality, 27% of total time; vocational problems, 14%; attitudes toward psychotherapy, 12%; attitudes toward and perception of self, 11%; society and the individual, 10%; interpersonal and social relations, 9%; relations to authority, 6%; family relations, 6%; handling of hostile feelings, 4%.

Roberts and Stroudbeck (78) applied the Bales system (Interaction Process Analysis) to group therapy interviews of depressive and schizophrenic patients. They expected

that the paranoid schizophrenic patients would use relatively more of the negative responses ("shows aggression" and "disagrees") and would make fewer responses to each other and more to the leader. The depressed patients were expected to make fewer responses per minute than the schizophrenic patients. All three of these predictions of differences between the groups were borne out by the data.

Gorlow *et al.* (28) studied nondirective group therapy sessions, using a system of categories similar to that developed by Snyder.

*Application of Bales system to student counseling.* Perry and Estes (69), with the collaboration of Bales, applied the Interaction Process Analysis system to four counseling interviews with a student. In this counseling case, the therapist shifted from responses categorized as "gives orientation" to responses falling in the categories "gives opinion" and "asks for opinion." The student showed a falling off, during the four interviews, in the percentage of responses classified as "shows tension, asks for help, withdraws out of the field." The authors interpreted the shift in therapist responses as documenting their belief that the counselor used non-directive techniques at the start of counseling, then shifted to responses that defined the counseling process as a collaborative venture. The drop in the client's "asks for help" responses was interpreted as showing his acceptance of a collaborative view of therapy and his giving up expectations that the therapist would play an authoritarian role.

*Studies of linguistic characteristics.* It is beyond the scope of this paper to discuss here studies of the grammatical structure of clients' speech during therapy. Pertinent studies include those of Roshal (86), Zimmer-

man and Langdon (99), and Grummon (29). These studies and others have been reviewed by Mowrer (60).

#### *Theoretically Guided Studies*

**Lasswell's pioneering work.** Lasswell was a pioneer not only in making sound recordings of analytic interviews, but also in applying content-analysis methods to their study. His earliest report on the analysis of transcribed interviews (45) indicates that he classified the client's utterances according to whether they referred to the interviewer or not. He proposed the hypothesis that *conscious affect* is indicated by references to the interviewer and that *unconscious tension* is indicated by slow speech, pauses, etc. If this hypothesis is correct, then tension as shown by slow speech should be correlated with other measures (e.g., physiological measures) of tension. In the cases studied Lasswell found a correlation between verbal indices of tension (slow speech) and physiological indices (high conductivity of the skin). On the other hand, conscious affect (indicated by references to interviewer) was correlated with increased heart rate. It is not clear to us why this latter relationship should have been found. Lasswell, however, interpreted it as supporting his hypothesis concerning the meaning of references to the therapist. The measures of unconscious tension (slow speech and high skin conductance) were negatively correlated with the measures of conscious affect (references to therapist and increased heart rate).

In another study (47) Lasswell obtained similar results. In this later study he included measures of blood pressure before and after each hour and of gross bodily movements during the hour. He found that unconscious affect (defined as above) decreased in the course of the inter-

views and conscious affect increased, as would be expected to occur in successful psychoanalytic treatment.

*Is clarification of feeling the best technique?* Rogers (82) has asserted his belief that progress in therapy always follows "recognition of feeling" by the therapist. Analytic therapists take a quite different view, believing that interpretation of resistance and labeling of previously unlabeled emotions are essential therapeutic techniques (Freud [23, 24], Dollard and Miller [20], Colby [12]). The available evidence from content-analysis studies concerning effectiveness of various techniques of the therapist is summarized in this section.

Snyder (91) studied the client responses following various kinds of responses by the therapist. He found that "insight" and "discussion of plans" by the client were more likely to follow nondirective than directive responses of the therapist.

Sherman (90) in a study of student counseling interviews found that "tentative analysis" by the therapist was most likely to be followed by client responses rated high on her "working relationship" scale; "interpretation" and "clarification" were moderately likely to be followed by good "working relationship;" and "urging" was the therapist technique least likely to be followed by good "working relationship." In discussing her study Robinson (79, p. 130) points out, however, that "a primary technique can have just about every degree of effect. Part of this range is due to the unreliability of rating scales, but in great part this range shows that at times each technique was a highly effective way of handling a particular unit or, in the case of poor outcomes, that the particular technique was the wrong one to use or that, with the interview conditions as they were, no

particular technique could have worked." Sherman also investigated the effect of technique used by the therapist on degree of insight shown by the client in the following response. In "adjustment-problem" units, clarification was most likely to be followed by a response rating high in "insight;" "interpretation" and "tentative analysis" were next most likely to be followed by "insight;" and "urging" was least likely to be so followed. It should be noted that the "insight" spoken of here is *insight-as-judged-by-Snyder's-or-Sherman's-judges*—and that other investigators might not have agreed with these authors concerning what is or is not insight. Psychoanalysts, for example, might require the patient to discover unconscious motivations before crediting him with the achievement of insight. Snyder and Sherman were dealing with something other than this.

In another study bearing on this point, Bergman (5) reached the conclusion that "reflection of feeling" was the only technique that led to insight or continued exploration. According to Bergman, interpretation by the therapist led to the abandonment of self-exploration. Contradictory results were obtained by Gillespie (27), who found that "verbal signs of resistance tend to be proportional to the number of counselor statements regardless of the counselor category" (27, p. 119). Hostile expressions toward the therapist and therapy were more likely to occur after interpretation or "inaccurate clarification of feeling" by therapist than after "restatement of content" and "accurate clarification of feeling." Other signs of "resistance" noted by Gillespie, however, were not significantly associated with interpretative activity by the therapist. These other signs in-

cluded: client-initiated long pauses, short answers, stereotyped repetition of the problem, changing the subject being discussed, excessive verbalization and intellectualizing, and emotional blocking. These signs of "resistance" are exactly those which psychoanalysts take as indicating resistance (12, pp. 96-98). Analytic therapists, however, apparently do not worry so much about the client's attacks and disagreements as the nondirective therapists do, except when these client responses threaten to interrupt communication.

In a study of a single case, Dittmann (18) noted what kinds of therapist responses were likely to be followed by client behavior that was judged to indicate progress. Client responses indicating "progress" were most likely to come after therapist responses that were slightly more interpretive than pure "reflection."

What can we learn from these studies? We learn, especially from the work of Robinson and his students, that there is no single type of response by the therapist that works best under all circumstances. Effectiveness of a response by the therapist probably depends on the expectations of the client (e.g., whether he expects the therapist to give advice or not), on the particular circumstances of the situation, and on the client's ability to tolerate increased self-knowledge at that particular time. We also learn that judgment about the effectiveness of a technique depends on having some measure of effect and that different measures of "effect" yield different judgments about what produces effectiveness. Measures of effectiveness (e.g., of "insight") that are considered appropriate by nondirective therapists may not necessarily be considered appropriate by analytic therapists—and vice versa.

*Differences in therapist responses.* Reid and Snyder (77) investigated the differences between counselors in their responses to identical client statements. Phonographically recorded client statements were re-played to a group of 15 clinical psychologists. After each statement the psychologists were allowed 15 seconds to write down the principal feelings they believed the client had expressed. Reid and Snyder found, first, that the same client response evoked a variety of responses from the counselors. They reported, further, that the counselors who were judged by their instructors to be more skillful were more likely to label the client's feeling in the way that a majority of the group labeled it. The authors also observed that some counselors consistently tended to label clients' feelings in a preferred way, e.g., some counselors were likely to perceive the clients as reporting feelings of insecurity, while other counselors labeled the same client responses as indicating hopefulness and ambition. This study shows a promising way to apply experimental methods to the study of the therapist's behavior.

*Haigh's study of defensiveness.* Haigh (31), believing that "defensiveness" (incorrect perception of one's behavior) should decrease during successful psychotherapy, identified instances of defensive behavior in 10 client-centered therapy cases, counting the number of these defensive responses in each interview. Haigh found that the clients had fewer defensive responses in the second half of treatment than in the first half. Seven of the clients had fewer defensive responses during the second half; two had more defensive responses during the second half; one had no identified defensive responses during either half. The

present authors have applied the exact test of significance to these data, and find that the first and second halves of therapy do not differ significantly ( $p=.30$ ).<sup>3</sup> A larger number of subjects must be studied before a definite conclusion concerning the course of defensiveness throughout therapy can be drawn.

A further aim of Haigh's study was to show that awareness of defensiveness caused a decrease in defensiveness. According to Haigh, the client would at the beginning of therapy present a number of defensive opinions. As the nondirective therapist reacted to the client with acceptance, lack of moral evaluation, and willingness to have the client make his own decisions, the client would have progressively weaker motives to present defensive opinions. He would become aware of inconsistencies in his opinions, actions, and emotions; in the benign atmosphere of therapy, he would be able to abandon these inconsistencies.<sup>4</sup> According to this account of therapy, increased awareness of defensiveness should be followed, in successful therapy, by a decrease in defensiveness. Haigh did not make the detailed analysis that would be necessary to test the adequacy of this construction, but he

<sup>3</sup> Haigh's statistical analysis, which resulted in a significant value of chi square, is inappropriate. The chi-square test requires the assumption that the items included in the table are independently selected. But the responses have not been independently selected from a population of responses; instead, if one response of a particular client was included in the sample, all his responses were included.

<sup>4</sup> Note the parallelism between this account of therapy and the account given by Dollard, Auld, and White (19). Dollard *et al.* believe, however, that the therapist must react with doubt and incredulity to elements of the client's account that do not make sense.

did report whether awareness-of-defensiveness showed any change, on the whole, throughout therapy. Haigh found that defensive behavior of which the client was unaware declined in the six cases that showed a drop in total defensive behavior; "unaware" defensive behavior increased in the three cases that showed an over-all increase in defensive behavior. Remembering that these results are derived from a very small sample of cases, we can make the tentative interpretation that clients in nondirective therapy who are unaware of their defensive behavior cannot change it; clients who become aware of it, can and do change it. We would like, however, to see the process of increased awareness studied in greater detail.

*A study of hostility and defenses.* An exploratory study making use of his content-analysis system has been reported by Murray (63). The hypotheses tested in this study were suggested by N. E. Miller's work on learning theory (55, 57, 58, 59, 67) and by psychoanalytic theory. Murray found that in the course of the psychotherapy case studied, defensive sentences of the client decreased and hostile sentences increased. He also found that interviews which contained a large number of defensive sentences had a relatively small number of hostile sentences ( $r = -.73$ ). Intellectualization and physical complaints comprised the defensive sentences; statements of frustration and resentment comprised the hostile sentences.

The results were interpreted in terms of learning theory as follows: The defenses were assumed to be motivated by anxiety. When the client uttered a defensive sentence instead of a hostile sentence, he thereby escaped the anxiety which would have been aroused by utter-

ance of the hostile sentence. During the psychotherapy the therapist interpreted one form of defense (physical complaints), while maintaining a permissive attitude toward the expression of hostility. These acts of the therapist resulted in a weakening of the defenses and extinction of the anxiety motivating them and thereby permitted an increased expression of hostility.

Murray also tested the hypothesis that the various hostile sentences would appear in a sequence which might be predicted by the theory of displacement (56, 58). His expectation was that the client would first express hostility about the less significant persons in his life and would gradually proceed to express hostility toward more important persons—for example, his mother. An examination of the data showed a sequence in the expression of hostility toward various persons, but not the sequence that had been predicted. Hostility was expressed first toward the more significant persons, then toward less significant persons. This finding led to a re-examination and reformulation of the displacement theory. Murray and Berkun made certain extensions of Miller's displacement theory and verified them in an animal experiment (66).

It was also discovered that the two defenses, intellectual discussion and physical complaints, seemed to operate as alternative members of a habit-family hierarchy. When one defense declined, the other tended to rise. Additional evidence for the hypothesis that the defenses functioned as members of a habit-family hierarchy is given by the fact that in the later stages of therapy, when anxiety was increasing following the greatly increased expression of hostility, the previously uninterpreted defense (intellectualization) in-



creased much more than the defense that had been interpreted.

*Attitudes toward self and others.*

Researchers of the nondirective group have investigated the changes in attitudes toward self and toward others that occur during psychotherapy. They believe that positive attitudes toward self and others are desirable, and they have been interested to discover what actions of the therapist can cause a change in these attitudes. One hypothesis is as follows: The therapist's warm, accepting attitude toward the client enables the client to recognize his own wishes more fully and evaluate his own behavior with greater acceptance (Rogers, 83, p. 41). A second hypothesis is this: To the degree that client has positive feelings about himself, he can have positive feelings toward other people. A third hypothesis is this: A person who feels affirmatively about other people gets along with them better than a person who reacts negatively. From these three hypotheses it follows that the therapist by a warm, accepting attitude toward the client can cause the client to feel positive about other people and to get along better with them (Rogers, 83, pp. 160, 520).

So far as we know, there has been no content study bearing on the first hypothesis. The second hypothesis was tested by Sheerer (89) and by Stock (93), who both found a correlation between positive opinions about self and positive opinions about others. The third hypothesis was investigated by McIntyre (52), who found no evidence that persons with high scores on an "acceptance-of-others" questionnaire were themselves accepted by others (as judged by a sociometric questionnaire). One must conclude that these three hypotheses have not yet been proved.

*Prognostic studies.* Can outcome

of therapy be predicted from early clues in the case? There seem to be only two content-analysis studies bearing on this question. Lasswell (46) investigated whether certain measures of speech and of physiological responses, made in early hours, could be used to predict the success of treatment. He reported that a combination of measures taken during the first 10 or 12 hours permitted accurate prediction of the client's progress. Lasswell reported of his subjects: "Those who adjusted themselves readily to the situation (and who made progress in insight) showed rising skin resistance curves. Our interpretation is that they were finding satisfactions in the interview experience against which they did not maintain a strong inner defense. Unconscious tension was diminishing, even though active conscious affect might rise" (46, pp. 246-247). Since the number of cases studied by Lasswell was small, this promising exploratory work needs to be repeated with larger samples.

The other study of prognostic indicators is that of Page (68), who found that measures of variability of the content and feeling of the client's speech in the first interview were not correlated with outcome of treatment. The amount of talk by the client in the first interview had a small ( $r = .30$ ) but statistically significant correlation with the criterion of success. It should be noted that Lasswell, Page, or any investigator who tries to find prognostic signs is hampered by the lack of any fully adequate definition or any adequate measure of "success" of therapy.

#### CHOICE OF A CONTENT-ANALYSIS SYSTEM

In choosing a content-analysis system, one faces all the problems of

choosing a topic that deserves scientific investigation *plus* the problems involved in choosing appropriate methods after one has decided on the topic. A few comments will be made on each of these tasks.

Experience has shown that the most fruitful scientific investigations are those that bear some relation to a general theory or, at least, to a well-thought-out hypothesis (Wilson, 98, p. 2). What hypotheses, then, might best be investigated? A large amount of attention has been focused on the immediate effects of various techniques used by the therapist and on the positive and negative opinions of patients. Little or no study by content-analysis methods has been made of such questions as these: What is the function of transference responses in psychotherapy? What is the effect of the therapist's giving love to the client? What is the role of childhood memories—must they be recaptured in successful therapy? How does the client learn new verbal units? What cues in the client's behavior does a good therapist respond to? These questions are relevant to a general theory of psychotherapy and thus, in our opinion, are worthy of study.

Once the investigator has chosen his field of inquiry, he must select appropriate techniques. If there are extant content-analysis systems which bear on his topic of inquiry, how can he decide whether to use one or more of them? This is the problem of validity. If the investigator is interested in the emotions of the client, he will want to know whether the D.R.Q. measures emotion. If he is interested in the influence of the therapist on the client, he may ask, "Does Porter's classification of therapist responses adequately measure this influence?" If he wants to assess therapeutic progress, he

might ask, "Do measures of 'self-attitude' such as Raimy's PNAvQ really reflect better adjustment of the client? Do these measures neglect unlabeled emotional conflicts? Do they give equal weight to changes in fundamental psychological conflicts and to resistant escape sentences?"

Janis (39) has proposed that the validity of a content system be judged by the number of relationships found by investigators who have used the system. As Janis points out, this does not mean that a valid measure is related to every other valid measure. For instance, intelligence may not be very important in determining the success of psychotherapy. If such be the case, we will not call the Wechsler-Bellevue Scale invalid if it fails to correlate with measures of success in therapy. But the intelligence-test score *should* be related to other measures of intelligence, e.g., to success in school work. Similarly, content measures are valid if they are correlated with other measures, when we have reason to believe they *should* be correlated. Admittedly, it is hard to designate which variables ought to be related to each other, and so it is difficult to determine when a lack of correlation implies lack of validity.

In our opinion, the content-analysis systems so far developed are not adequate to the task of marking out the main variables in therapy. Most of them rely too much on the opinions presented by clients and neglect clients' unconscious motives. Some investigators have assumed that an increase in favorable opinions about himself indicates improvement in the client. While we believe that this is sometimes true, we are aware that clients can change their opinions in response to stimuli other than im-

proved psychological well-being. For instance, a client having an unconscious motive to escape the anxiety evoked by therapy may offer the opinion that he feels much better and has been enormously helped. Therefore, the client says, he is ready to quit psychotherapy. Again, a client motivated by the desire to please the therapist may say that he has been greatly benefited. He would feel ungrateful if he did not acknowledge benefit. These two motives, used as illustrations, are only two of many possible motives, both conscious and unconscious, that can influence the client's expression of an opinion about himself. Our knowledge that such motivations exist warns us not to rely too much on these opinions.<sup>8</sup>

Content systems are inevitably criticized for what they leave out. The practicing clinician often feels that the measured part of the therapeutic transaction is pitifully small alongside the complex of stimuli that he senses as a participant observer. Yet it seems unfair to expect any single content-analysis system to describe all of this complex situation. We would probably make a fairer appraisal of content systems if we expected each system to deal with only a part of this complexity. An adequate descriptive and causal an-

alysis of psychotherapy will most likely require a large number of measures, each of them shown to be reliable and valid for its limited purpose. Measures of the content of clients' and therapists' utterances will undoubtedly be supplemented by measures of other, nonverbal responses of client and therapist. By the combination of a variety of measures, each useful in its own domain, we may in time construct an adequate science of psychotherapy.

#### SUMMARY

Research on psychotherapy has been hampered until recently by the lack of permanent records of the transaction, by absence of objective measures, and by lack of an appropriate theoretical framework. The advent of sound recording of interviews, the widespread application of content-analysis methods, and the development of psychological theories such as learning theory have opened up new possibilities for research on psychotherapy. This paper surveys some of the first fruits of the new developments.

Studies are here classified as methodological (development of measures), descriptive, or theoretically oriented. An attempt has been made to include representative and important studies of each type. It is believed that the value of studies in this field must finally be assessed in terms of relevance to theory. Studies that only result in the development of new measuring instruments or in the publication of a description of some type of therapy are less valuable than studies that test some hypothesis. The methodological and descriptive investigations may be valuable, however, if they provide the tools and the hypotheses that make later theoretically guided investigation possible.

<sup>8</sup> We believe that an objective system of content analysis need not neglect unconscious motives. The therapist who attributes unconscious motives to his patient is, when he does this, reacting to cues provided by the patient's behavior. It should be possible to teach other persons to react similarly to the patient's behavior. If the responses of the patient that reveal unconscious motivation are verbal, it would be possible to designate these unconscious reactions by the content-analysis method. For example, if overconcern about another person's health indicates unconscious hostility toward that person, any sentence expressing overconcern can be scored as "unconscious hostility."

## REFERENCES

1. ARONSON, M. A study of the relationships between certain counselor and client characteristics in client-centered therapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Coll., 1953. Pp. 39-54.
2. ASSUM, A. L., & LEVY, S. J. Analysis of a nondirective case with followup interview. *J. abnorm. soc. Psychol.*, 1948, 43, 78-89.
3. BALES, R. F. *Interaction process analysis*. Cambridge: Addison-Wesley, 1950.
4. BERELSON, B. *Content analysis in communication research*. Glencoe, Ill.: Free Press, 1952.
5. BERGMAN, D. Counseling method and client responses. *J. consult. Psychol.*, 1951, 15, 216-224.
6. BLAU, B. A. A comparison of more improved with less improved clients treated by client-centered methods. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Coll., 1953. Pp. 120-126.
7. BRODY, E. B., NEWMAN, R., & REDLICH, F. C. Sound recording: the problem of evidence in psychiatry. *Science*, 1951, 113, 379-380.
8. BUGENTAL, J. F. T. A method for assessing self and not-self attitudes during the therapeutic series. *J. consult. Psychol.*, 1952, 16, 435-439.
9. CARNES, E. F., & ROBINSON, F. P. The role of client talk in the counseling interview. *Educ. psychol. Measmt.*, 1948, 8, 635-644.
10. COFER, C. N., & CHANCE, JUNE. The discomfort-relief quotient in published cases of counseling and psychotherapy. *J. Psychol.*, 1950, 29, 219-224.
11. COFFEY, H. S., FREEDMAN, M., LEARY, T., & OSSORIO, A. Community service and social research—group psychotherapy in a church program. *J. soc. Issues*, 1950, 6, 1-65.
12. COLBY, K. M. *A primer for psychotherapists*. New York: Ronald Press, 1951.
13. COLLIER, R. M. A scale for rating the responses of the psychotherapist. *J. consult. Psychol.*, 1953, 17, 321-326.
14. COVNER, B. J. Studies in phonographic recordings of verbal material: I. The use of phonographic recordings in counseling practice and research. *J. consult. Psychol.*, 1942, 6, 105-113.
15. COVNER, B. J. Studies in phonographic recordings of verbal material: III. The completeness and accuracy of counseling interview reports. *J. gen. Psychol.*, 1944, 30, 181-203.
16. COVNER, B. J. Studies in phonographic recordings of verbal material: IV. Written reports of interviews. *J. appl. Psychol.*, 1944, 28, 89-98.
17. CURRAN, C. A. *Personality factors in counseling*. New York: Grune & Stratton, 1945.
18. DITTMANN, A. T. The interpersonal process in psychotherapy: development of a research method. *J. abnorm. soc. Psychol.*, 1952, 47, 236-244.
19. DOLLARD, J., AULD, F., & WHITE, ALICE M. *Steps in psychotherapy*. New York: Macmillan, 1953.
20. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
21. DOLLARD, J., & MOWLER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, 42, 3-32.
22. ELTON, C. F. A study of client responsibility: counselor technique or interview outcome? *Educ. psychol. Measmt.*, 1950, 10, 728-737.
23. FREUD, S. Further recommendations in the technique of psychoanalysis. On beginning the treatment. The question of the first communications. The dynamics of the cure. In *Collected papers*, Vol. 2. London: Hogarth and the Institute of Psychoanalysis, 1924. Pp. 342-365.
24. FREUD, S. Further recommendations in the technique of psychoanalysis. Recollection, repetition, and working through. In *Collected papers*, Vol. 2. London: Hogarth and the Institute of Psychoanalysis, 1924. Pp. 366-376.
25. GELLER, A., KAPLAN, D., & LASSWELL, H. D. An experimental comparison of four ways of coding editorial content. *Journ. Quart.*, 1942, 19, 362-370.
26. GILL, M., NEWMAN, R., & REDLICH, F. C. *The initial interview in psychiatric practice*. New York: International Universities Press, 1954.
27. GILLESPIE, J. F. Verbal signs of resistance in client-centered therapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Coll., 1953. Pp. 105-119.
28. GORLOW, L., HOCH, E. L., & TILSCHOW, E. F. *The nature of nondirective group psychotherapy: an experimental investi-*

- gation. New York: Teachers Coll., Columbia Univer., 1952.
29. GRUMMON, D. L. An investigation into the use of grammatical and psycho-grammatical categories of language for the study of personality and psychotherapy. Unpublished doctor's dissertation, Univer. of Chicago, 1950. (Not seen; cited by Mowrer [60].)
  30. GUMP, P. V. A statistical investigation of one psychoanalytic approach and a comparison of it with non-directive therapy. Unpublished master's thesis, Ohio State Univer., 1944.
  31. HAIGH, G. Defensive behavior in client-centered therapy. *J. consult. Psychol.*, 1949, 13, 181-189.
  32. HILGARD, E. R. Experimental approaches to psychoanalysis. In E. Pumpian-Mindlin (Ed.), *Psychoanalysis as science*. Stanford, Calif.: Stanford Univer. Press, 1952. Pp. 3-45.
  33. HOFFMAN, A. E. Reported behavior changes in counseling. *J. consult. Psychol.*, 1949, 13, 190-195.
  34. HOGAN, R. A. A measure of client defensiveness. In W. Wolff and J. A. Precker (Eds.), *Success in psychotherapy*. New York: Grune & Stratton, 1952. Pp. 112-142.
  35. HORWITZ, FLORENCE S. The relationship between a method of measuring tension and two methods of measuring the attitude toward self in thirty-six electrically recorded initial social casework interviews. M.S. thesis, Boston Univer. Sch. of Soc. Work, 1951.
  36. HUNT, J. McV. Measuring the effects of social casework. *Trans. N. Y. Acad. Sci.*, 1947, 9, 78-88.
  37. HUNT, J. McV. Measuring movement in casework. *J. soc. Casework*, 1948, 29, 343-351.
  38. HUNT, J. McV. A social agency as a setting for research—the Institute of Welfare Research. *J. consult. Psychol.*, 1949, 13, 69-81.
  39. JANIS, I. L. The problems of validating content analysis. In H. D. Lasswell, N. Leites, et al., *Language of politics*. New York: Stewart, 1949. Pp. 55-82.
  40. JONES, E. *The life and work of Sigmund Freud*, Vol. 1. New York: Basic Books, 1953.
  41. KAHN, M. W. The role of perceptual consistency and generalization change in Rorschach and psychotherapy behavior. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Coll., 1953. Pp. 115-134.
  42. KAPLAN, A., & GOLDSSEN, J. M. The reliability of content-analysis categories. In H. D. Lasswell, N. Leites, et al., *Language of politics*. New York: Stewart, 1949. Pp. 83-112.
  43. KAUFFMAN, P. E., & RAIMY, V. C. Two methods of assessing therapeutic progress. *J. abnorm. soc. Psychol.*, 1949, 44, 379-385.
  44. KUBIE, L. S. Problems in clinical research. *Amer. J. Orthopsychiat.*, 1947, 17, 196-203.
  45. LASSWELL, H. D. Verbal references and physiological changes during the psychoanalytic interview: a preliminary communication. *Psychoanal. Rev.*, 1935, 22, 10-24.
  46. LASSWELL, H. D. Certain prognostic changes during trial (psychoanalytic) interviews. *Psychoanal. Rev.*, 1936, 23, 241-247.
  47. LASSWELL, H. D. Veränderungen an einer Versuchsperson während einer Kurzen Folge von psychoanalytischen Interviews. *Imago*, 1937, 23, 375-380.
  48. LASSWELL, H. D. A provisional classification of symbol data. *Psychiatry*, 1938, 1, 197-204.
  49. LASSWELL, H. D., LEITES, N., et al. *Language of politics*. New York: Stewart, 1949.
  50. LEWIS, VIRGINIA W. Changing the behavior of adolescent girls. *Arch. Psychol.*, 1943, No. 279.
  51. LINDNER, R. M. *Rebel without a cause*. New York: Grune & Stratton, 1944.
  52. MCINTYRE, C. J. Acceptance by others and its relation to acceptance of self and others. *J. abnorm. soc. Psychol.*, 1952, 47, 624-625.
  53. MEADOW, A., GREENBLATT, M., LEVINE, J., & SOLOMON, H. C. The Discomfort-Relief Quotient as a measure of tension and adjustment. *J. abnorm. soc. Psychol.*, 1952, 47, 658-661.
  54. MILLER, HELEN E. "Acceptance" and related attitudes as demonstrated in psychotherapeutic interviews. *J. clin. Psychol.*, 1949, 5, 83-87.
  55. MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald Press, 1944. Pp. 431-465.
  56. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, 43, 155-178.
  57. MILLER, N. E. Comments on theoretical models, illustrated by the development



- of a theory of conflict behavior. *J. Pers.*, 1951, 20, 82-100.
58. MILLER, N. E., & KRAELING, DORIS. Displacement: greater generalization of approach than avoidance in a generalized approach-avoidance conflict. *J. exp. Psychol.*, 1952, 43, 217-221.
  59. MILLER, N. E., & MURRAY, E. J. Displacement and conflict: learnable drive as a basis for the steeper gradient of avoidance than of approach. *J. exp. Psychol.*, 1952, 43, 227-231.
  60. MOWRER, O. H. Changes in verbal behavior during psychotherapy. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research*. New York: Ronald Press, 1953. Pp. 463-545.
  61. MOWRER, O. H., HUNT, J. McV., & KOGAN, L. S. Further studies utilizing the Discomfort-Relief Quotient. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research*. New York: Ronald Press, 1953. Pp. 257-295.
  62. MOWRER, O. H., LIGHT, B. H., LURIA, ZELLA, & ZELNY, MARJORIE P. Tension changes during psychotherapy with special reference to resistance. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research*. New York: Ronald Press, 1953. Pp. 546-640.
  63. MURRAY, E. J. A case study in a behavioral analysis of psychotherapy. *J. abnorm. soc. Psychol.*, 1954, 49, 305-310.
  64. MURRAY, E. J. A method for studying psychotherapy. Unpublished doctor's dissertation, Yale Univer., 1955.
  65. MURRAY, E. J., AULD, F., JR., & WHITE, ALICE M. A psychotherapy case showing progress but no decrease in the Discomfort-Relief Quotient. *J. consult. Psychol.*, 1954, 18, 349-353.
  66. MURRAY, E. J., & BERKUN, M. M. Displacement as a resolution of conflict. *J. abnorm. soc. Psychol.*, 1955, 1, 47-56.
  67. MURRAY, E. J., & MILLER, N. E. Displacement: steeper gradient of generalization of avoidance than of approach with age of habit controlled. *J. exp. Psychol.*, 1952, 43, 222-226.
  68. PAGE, H. A. An assessment of the predictive value of certain language measures in psychotherapeutic counseling. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Coll., 1953. Pp. 88-93.
  69. PERRY, W. G., & ESTES, S. G. The collaboration of client and counselor. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research*. New York: Ronald Press, 1953. Pp. 95-119.
  70. PORTER, E. H. The development and evaluation of a measure of counseling interview procedures. I. The development. *Educ. psychol. Measmt*, 1943, 3, 105-126.
  71. PORTER, E. H. The development and evaluation of a measure of counseling interview procedures. II. The evaluation. *Educ. psychol. Measmt*, 1943, 3, 215-238.
  72. PORTER, E. H. *An introduction to therapeutic counseling*. Boston: Houghton Mifflin, 1950.
  73. RAIMY, V. C. Self reference in counseling interviews. *J. consult. Psychol.*, 1948, 12, 153-163.
  74. RAKUSIN, J. M. The role of Rorschach variability in the prediction of client behavior during psychotherapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Coll., 1953. Pp. 60-74.
  75. RASKIN, N. J. An objective study of the locus-of-evaluation factor in psychotherapy. In W. Wolff and J. A. Precker (Eds.), *Success in psychotherapy*. New York: Grune & Stratton, 1952. Pp. 143-162.
  76. REDLICH, F. C., DOLLARD, J., & NEWMAN, R. High fidelity recording of psychotherapeutic interviews. *Amer. J. Psychiat.*, 1950, 107, 42-48.
  77. REID, DOROTHY, & SNYDER, W. U. Experiment in "recognition of feeling" in non-directive psychotherapy. *J. clin. Psychol.*, 1947, 3, 128-135.
  78. ROBERTS, B. H., & STRODTBECK, F. L. Interaction process differences between groups of paranoid schizophrenic and depressed patients. *Int. J. group Psychother.*, 1953, 3, 29-41.
  79. ROBINSON, F. P. *Principles and procedures in student counseling*. New York: Harper, 1950.
  80. ROGERS, C. R. *Counseling and psychotherapy*. Boston: Houghton Mifflin, 1942.
  81. ROGERS, C. R. Electrically recorded interviews in improving psychotherapeutic techniques. *Amer. J. Orthopsychiat.*, 1942, 12, 429-435.
  82. ROGERS, C. R. The development of insight in a counseling relationship. *J. consult. Psychol.*, 1944, 8, 331-341.
  83. ROGERS, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
  84. ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.

85. ROSENMAN, S. Changes in interpersonal relations in the course of non-directive counseling. Unpublished doctor's dissertation, Harvard Univer., 1953.
86. ROSHAL, JEAN J. G. The type-token ratio as a measure of changes in behavior variability during psychotherapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Coll., 1953. Pp. 94-104.
87. ROYER, ANNE E. An analysis of counseling procedures in a non-directive approach. Unpublished master's thesis, Ohio State Univer., 1943. (Not seen; cited by Snyder [91].)
88. SEEMAN, J. The process of non-directive therapy. *J. consult. Psychol.*, 1949, 13, 157-168.
89. SHEERER, ELIZABETH T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, 13, 169-175.
90. SHERMAN, DOROTHY. An analysis of the dynamic relationship between counselor techniques and outcomes in larger units of the interview situation. Unpublished doctor's dissertation, Ohio State Univer., 1945. (Not seen; cited by Robinson [79].)
91. SNYDER, W. U. An investigation of the nature of non-directive psychotherapy. *J. gen. Psychol.*, 1945, 33, 193-223.
92. SNYDER, W. U. A comparison of one unsuccessful with four successful non-directively counseled cases. *J. consult. Psychol.*, 1947, 11, 38-42.
93. STOCK, DOROTHY. The self concept and feelings toward others. *J. consult. Psychol.*, 1949, 13, 176-180.
94. STRUPP, H. H. An objective comparison of Rogerian and psychoanalytic techniques. *J. consult. Psychol.*, 1955, 19, 1-7.
95. STRUPP, H. H. An objective study of certain psychotherapeutic operations; the effect of the therapist's personal analysis upon his verbal techniques. *Amer. Psychologist*, 1954, 9, 479-480. (Abstract)
96. TUCKER, J. E. Measuring client progress in client-centered therapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State Coll., 1953. Pp. 55-59.
97. WHITE, R. K. *Black boy*: A value analysis. *J. abnorm. soc. Psychol.*, 1947, 42, 440-461.
98. WILSON, E. B., JR. *An introduction to scientific research*. New York: McGraw-Hill, 1952.
99. ZIMMERMAN, W., & LANGDON, J. A preliminary attempt to establish criteria for measuring progress in psychotherapy. Unpublished manuscript, Univer. of Illinois. (Not seen; cited by Mowrer [60].)

Received July 7, 1954.

## EMPLOYEE ATTITUDES AND EMPLOYEE PERFORMANCE<sup>1</sup>

ARTHUR H. BRAYFIELD AND WALTER H. CROCKETT

*Kansas State College*

The systematic investigation of employee attitudes is a relatively recent development in American business and industry. Although Houser and his associates (26) pioneered in this field in the early 1920's there was little active interest until early in World War II when employee attitude surveys began to flourish (49, p. 7). Currently there is an abundant and growing literature on the use of this personnel tool (56).

Only infrequently, however, are discussions of the correlates of employee attitudes found and these are almost never substantiated by empirical evidence. Where we have located relevant discussions in the personnel and psychological literature a common assumption predominates—employee attitudes bear a significant relationship to employee performance. These are sample quotations: "... morale is not an abstraction. Rather it is concrete in the sense that it directly affects the quality and quantity of an individual's output." "Numerous investigations have established the certainty that productive efficiency fluctuates with variations in interest and morale." "... employee morale ... reduces turnover. It makes labor trouble and strikes less likely. It cuts down absenteeism and tardiness; lifts production."

It is of some practical and theoretical interest to establish the relationships which exist between employee

attitudes and employee performance. The purpose of this review is to examine and summarize the empirical literature which bears upon these relationships and to engage in some discussion of the methodological and theoretical considerations involved in such investigations.

Examination of the literature reveals that it is (a) recent, and (b) frequently peripheral in the sense that relevant data were collected and analyzed incidental to some other objective.

We have established certain conditions for the inclusion of materials in this review. First, the indices of employee attitudes must permit classification of respondents along some attitude continuum. Second, the indices of employee attitudes must have been obtained directly from the employees themselves. Although we are willing to include ratings of job performance by supervisors and others, if no other criteria of performance are available, we are not willing to accept estimates of *attitudes* by someone other than the individuals themselves. Performances, we would contend, are less easily disguised by the individual and less readily distorted by the observer than are attitudes. Third, the investigations must have been conducted in industrial or occupational settings. Within the limitations of interlibrary loan service our coverage is complete through July, 1954. We have made no effort to unearth unpublished studies although we report several including three studies by one of us.

<sup>1</sup> Portions of this paper were presented by Brayfield before the University of Minnesota Chapter of Psi Chi in March, 1954.

The following scheme was adopted as a convenient and meaningful way of categorizing the literature.

1. Daniel Katz and Robert Kahn (33, p. 657) have suggested that "in social structures it is important to distinguish between: (1) the motivation to stay within the system, to remain a part of the group and (2) the motivation to act in a differential manner within that system." We have thus distinguished between those studies which involve performance on the job and those which involve withdrawal from the job (absences, accidents, turnover).

2. Within the above breakdown we have made a further differentiation based upon research design. One major design relates the attitudes of individuals to their performances as individuals. A second design relates the attitudes of the members of groups to their performances as groups.

3. A still further classification differentiates between studies in which a single index of attitudes either as a single item or as a summation of items was used and those few in which multiple indices were used.

We have not attempted to define such terms as job satisfaction or morale. Instead, we have found it necessary to assume that the measuring operations define the variables involved. Definitions are conspicuous by their absence in most current work in this area.

Where reliability data are reported for the attitude and performance measures, we have included them in our summaries. We also have attempted to specify whether or not the attitude data were collected under conditions which preserved the anonymity of the subjects. Throughout the first section of the paper we have tried to hold comments on methodology to a minimum, postponing de-

tailed methodological considerations until the substantive material has been covered.

Before summarizing and discussing the literature it may be appropriate to describe the investigation which, as far as we can determine, initiated research in this area of industrial psychology. The classic study relating attitudes and performance in an industrial setting was conducted by Kornhauser and Sharp (39) in 1930 in Neenah, Wisconsin, in the mill operated by the Kimberly-Clark Corporation. Between 200 and 300 young girls engaged in routine repetitive jobs at machines were studied. Both questionnaires and interviews were used. The questionnaires were patterned after those developed by Houser and covered a range of specific attitudes—toward supervisors, repetitiveness and speed of work, personnel policies, wages, and the like. Scores were computed for groups of items and item responses were analyzed. Intercorrelations among different item groups ran about .4 to .5. Reliabilities were thought to be somewhat higher.

The finding on relationship of attitudes to performance is summed up in the statement that "Efficiency ratings of employees showed no relationship to their attitudes." No description is given of the rating system. Further, the authors say, "In one group of 20 girls for whom we had comparable output records, three of the four with the most unfavorable attitudes were first, second, and fourth in production and the two most favorable were near the bottom in production."

With respect to the criterion of withdrawal from the job, Kornhauser and Sharp reported that "Unfavorableness of job attitudes is slightly correlated with lost time because of sickness."

Relations between attitudes and intelligence, age, schooling, marital status, home life, emotional adjustment, and supervision also were studied. This early report should be read by anyone seriously interested in this area of investigation.

#### PERFORMANCE ON THE JOB

##### *Individual Analysis*

Three unpublished studies have used the Brayfield-Rothe Job Satisfaction Blank as an index of job satisfaction. In 1943 Brayfield (4) started work on the development of a scale intended to give what might be called a global measure of job satisfaction. It was predicated on attitude theory and applied the Thurstone scaling technique. After some preliminary work Likert's scoring technique was applied to 18 Thurstone-scaled items to produce an index which had a range of scores from 18 through 90 with a neutral or indifferent point at 54. The resulting scale gave a corrected split-half reliability coefficient of .87 when used with 231 women office employees. It differentiated between adults enrolled in a night class in personnel psychology who were employed in personnel jobs and similar students who were employed in non-personnel jobs. For the same group, a correlation of .91 was obtained between the Brayfield-Rothe and the Hoppock Job Satisfaction Blank.

In 1944, in connection with a larger study Brayfield collected data on 231 women office employees working for the same firm but employed in 22 different offices throughout the country. The scale was administered to small groups of individuals as part of a test battery. All materials were signed. At the same time supervisor's ratings on a graphic rating scale were obtained for all employees in the sample. A total score was computed from

three items covering quantity, quality, and over-all worth to the company. About two-thirds of the employees were rated by two supervisors and the ratings were averaged. Reliabilities of ratings are unknown although in one office it was possible to compare two supervisors who had rated the same 23 women. The inter-correlation was in the low seventies.

When job satisfaction scores for these women clerical workers were compared with their performance ratings a correlation of .138, significant at the 5% level, was found. To control for the influence of job level, the 231 women were classified into six groups as follows: Stenographers (50); General Clerical (40); Typists (38); High Level Machine Clerical (36); Low Level Machine Clerical (34); Entry (33). The correlations for the first five groups ranged from -.06 to +.13. None were significant. The correlation for the group of 33 inexperienced and untrained girls (Entry) was .387, significant at the 5% level. An additional group of 35 women telephone order clerks provided a correlation of .26 which was not significant.

In 1950, Brayfield and Mangelsdorf obtained data on 55 second-, third-, and fourth-year plumber apprentices employed in a number of firms in Oakland, California. All were enrolled four hours per week in a public vocational school. The subjects completed the Brayfield-Rothe job satisfaction scale during classes as part of a testing program in which all the materials were identified by name of respondent. The corrected split-half reliability coefficient was .83. Performance ratings were obtained for each plumber from his foreman or employer. The rating form consisted of 25 scaled items in check list form developed by Goertzel (19, p. 117) who attempted to provide



a generalized scale that could be used for assessment of workers on any type of job. For various groups of workers Goertzel found a correlation of approximately .80 between ratings on two forms of 25 items each. The correlation between job satisfaction scores and ratings was .203 which is not significant.

In 1953, Brayfield and Marsh studied the measured characteristics of 50 farmers enrolled four hours per week in a veterans' on-job training program. The median age of the subjects was in the early thirties. They had lived on farms most of their lives; all were managing their own farms. Among other materials they completed the Brayfield-Rothe job satisfaction scale. All materials were signed. The corrected split-half reliability coefficient was .60; if the three subjects with the most inconsistent responses had been eliminated, the reliability coefficient would have become .77.

The subjects' performance as farmers was rated by their instructors who ranked them in order of effectiveness. Sixteen farmers were ranked by one instructor, 14 by a second, and the remaining 20 by another. Ranks were transmuted into "scores" (18) and cast into a single distribution. Re-rankings after several months, when similarly treated, correlated .86 with the original rankings. Instructors were not aware that they would be asked to re-rank their students.

For the 50 farmers the correlation between job satisfaction scores and performance ratings was .115 which is not significant. If the three "erratic" subjects had been eliminated, the correlation would have become .133.

The same job satisfaction scale was used in 1953 in an unpublished study by Roger Bellows and associates of 109 Air Force control tower opera-

tors.<sup>2</sup> The correlation with individual proficiency ratings was .005.

Gadel and Kriedt (17) report a study employing a design similar to that used in the investigations just described. One hundred and ninety-three male IBM operators working in the machine rooms of numerous divisions of the Prudential Insurance Company home office completed and signed a 10-item job satisfaction questionnaire "designed to cover a variety of attitudes related to work duties." The performance criterion consisted of rank-order ratings on over-all job performance made by the immediate supervisor. Ratings were converted to standard scores and correlations were computed for each of the groups. The resulting correlations were averaged using the Fisher  $z$  transformation. The relationship between job satisfaction and performance was found to be .08.

The Life Insurance Agency Management Association has engaged in job satisfaction studies since the early 1940's. A report of their work which falls into the research classification under consideration was published by Habbe (23) in 1947. Job satisfaction questionnaires were mailed out to 9,353 insurance agents. Seventy-five per cent were returned of which more than 90% were usable. Signatures were not requested although quite a few agents did identify themselves. The blank contained questions asking about single phases of the job to be answered by one of five alternatives indicating satisfaction or dissatisfaction. A single question asked "How do you feel about your job as a life underwriter?" The performance rating was in the form of a self-report since each agent was asked to check whether his previous year's production was "under

<sup>2</sup> R. M. Bellows, Personal communication, June 30, 1954.

\$200,000" of insurance or \$200,000 or over." Agents producing under \$200,000 scored 4.15 on what evidently was the single general satisfaction item as compared to 4.11 for the high producers. The "Extremely Satisfied" score is 5.00. The relationship is insignificant or slightly in favor of the lower producers. It should be noted that this performance criterion is a self-report and that the break at the \$200,000 point might not be the best point for analyzing the relationship.

Baxter and his associates (1) have recently reported a training evaluation study concerned with new debit insurance agents (service and sell weekly and monthly premium, ordinary, and group insurance for families within a specific geographical territory). Included in the data collected were responses to a comprehensive job satisfaction attitude questionnaire with items varying in number from 32 to 43 depending upon when it was administered. Respondents apparently were identified. Supervisor ratings on a 5-point, 9-item graphic rating scale were collected. Sales volume figures were obtained for each agent for his first year on the job. Although the correlations between the job satisfaction index and the performance criteria were not reported, the investigators have made them available.<sup>3</sup> For 223 agents the correlation between satisfaction and supervisor's rating is .23, significant at the 1% level. The correlation between satisfaction and sales volume is .26 also significant at the 1% level. This is the only study in this classification which uses an objective performance criterion. The incentive situation is also more clear-cut here except perhaps for the farmers. Al-

though this correlation is significant, it is quite low.

One of the most carefully done studies which we have inspected is Mossin's (48) investigation of the selling performance and what he termed contentment of 94 teen-age female retail sales clerks in a large New York department store. His performance criteria of 12 items were based on the ratings of four experienced and specially trained shoppers. Ratings on five items formed a composite labeled "Selling Attitudes." Ratings on three other items were combined as an index of "Selling Skills." The intercorrelation was .76. In addition, the entire 12 items formed a composite which correlated beyond .9 with each of the other criteria. Several detailed analyses of the reliability of the criteria were made including intercorrelations among the four shoppers. A minimum estimate of the reliability of the criteria would be that they exceeded .7 and might actually have been somewhat higher.

Mossin used two job satisfaction measures. One was an over-all composite rating secured by combining the scores on 6 attitude items inquiring about "affective dispositions" toward departmental assignment, merchandise assignment, relations with customers, relations with fellow salesgirls, relations with supervisors, and working conditions, along with one item regarding intention to remain in retail selling plus one item requiring a self-appraisal of sales ability. The second index was a single multiple-response item asking "How you REALLY feel toward your job." Responses on these two indices of job satisfaction were obtained during an individual data collection session with the investigator. Therefore, the respondents were identified. The correlation between the two satisfaction criteria was .53.

<sup>3</sup> B. Baxter. Personal communication, February 17, 1954.

The composite job satisfaction score correlated  $-.07$  with the Attitudes criterion and  $-.03$  with the Skill criterion. The single item job satisfaction index correlated  $.15$  and  $.06$ , respectively. None of these is significant. No results were reported for the 12-item composite performance criterion although it may be inferred that they would be of approximately the same magnitude since it was highly correlated with the other two criteria. This is a carefully executed investigation and should be consulted by anyone working in this general area.

The final major investigation in this series, by Bernberg (2), is the only one to use differentiated attitude measures. He included a measure of group morale as identified by 34 "indirect method" items, a 12-item scale presumed to measure an employee's acceptance of the formal organization (e. g., "I think this company treats its employees worse than any other company does"), a 0-100 thermometer scale with seven verbal referent points based on the single statement, "On the whole, I believe that the supervisor in my group is a man who knows his job and is a leader," and a similar thermometer scale for the self-rating statement, "On the whole, I believe that my group has a high degree of morale. By that, I mean the men work willingly and cheerfully as a well organized team." The intercorrelations among the four indices as computed for 890 hourly paid workers in a large aircraft manufacturing plant ranged from  $.47$  to  $.77$  with the median at  $.5$ . Split-half reliabilities for the two multi-item scales were approximately  $.8$ .

Questionnaires embodying these measures were sent home with more than 1,000 employees of an aircraft plant. No returns were accepted after 48 hours by which time 88% were

back. Presumably the respondents were identified. The performance criterion was the average weighted score of a graphic rating scale with the five dimensions of adaptability, dependability, job knowledge, quality, quantity. The split-half reliability was  $.8$ .

The correlations between the four attitude measures and the performance criterion ranged from  $.02$  to  $.05$ .

Four miscellaneous studies warrant brief mention only. An English doctoral dissertation (40) is reported to include the finding that it was clearly determined that "there is almost no relationship between proficiency and satisfaction among (British) post office counter clerks." Kerr (38) reports a master's study finding of a correlation of  $-.76$  between a 10-item job satisfaction measure and employer reports on the frequency of what he termed grievance, advice, and catharsis conferences with employees in two very small Indiana plants. The study is relevant mainly as suggesting a possible performance criterion for investigation. Chase (7) has a very inadequately described study which purports to find a small positive relationship between superintendent's ratings and teacher's satisfaction. Brody's (5) master's thesis at New York University describes an investigation in which the relationship between Hoppock Job Satisfaction Blank scores and production under a piece work incentive plan correlated  $.68$  for 40 employees. This is an extraordinary finding. However, examination of the raw data in the Appendix casts serious doubt on the meaningfulness of the correlation. Two groups working under different incentive conditions are lumped together. For the 22 cases which might

\* Reported by Heron, A. Industrial psychology. *Ann. Rev. Psychol.*, 1954, 5, 203-228.

actually be legitimate subjects, the Hoppock scores do not conform to any known appropriate scoring system for that particular Blank. The production data are bimodal.

At this point we can summarize the findings for this research design. The prototype study used a single over-all index of employee attitudes variously titled job satisfaction or morale. Respondents were identified. A distribution of individual scores was related to some index of individual performance on the job. Customarily, a single occupational group was studied. When 14 homogeneous occupational groups and one large sample of assorted hourly factory workers were studied, statistically significant low positive relationships between job satisfaction and job performance were found in two of the 15 comparisons. These results, pointing to an absence of relationships, are in line with the findings of the pioneering Kornhauser and Sharp investigation.

#### *Group Analysis*

The essentials of this design are as follows. Employee attitudes are determined individually but the average for the group or the percentage responding in a certain manner is related to some estimate of performance or productivity for the group as a whole. This arrangement requires at least two groups. Characteristically, comparisons are by departments within a firm rather than by occupation.

The antecedents of this approach are to be found in a study by Rensis Likert which was reported in a privately circulated document in 1941. We have not examined the report. According to a reference to it in one of Katz's (30) papers, the morale of insurance agents in 10 agencies rated superior in operational efficiency by

the home offices of nine companies was compared to that of agents in 20 agencies rated below average. We infer that some form of attitude questionnaire was used since Likert conducted the study although interviews may have been involved. Katz says that "Morale was found to be significantly related to the criterion." This study is mainly of historical interest.

Three studies employing this design or a modification of it utilized a single index of employee attitudes. Katz and Hyman (32) report a study which they supervised during World War II under the general direction of Likert. Their concern was with employee morale in shipyards and its relation to productivity, among other considerations. Two summary measures of morale were used, both of which were obtained from personal interview protocols. One was a yes-no answer to the question: "Have you ever felt like quitting the yards?" The rank order of the percentage who had felt like quitting was compared with an index of productivity (time to turn out a ship) for the five shipyards being studied. The two rank orders agreed fairly well. The second measure of employee attitudes was furnished by the responses to 7 items regarding specific aspects of the working situation and environment. The relationship to productivity was somewhat less marked than the first comparison although the authors comment that "In general the yards with high productivity were the yards with high worker morale." It should be remarked that, although the productivity differences were very great among the yards, the morale differences were really quite small; the morale scores for the five yards were 9.3, 9.4, 10.0, 10.0, 10.9.

Giese and Ruter (20) employed this design in a study of employees

of a small national mail order company. It is one of the few studies which had as its primary purpose the determination of the relationship in which we are interested. In fact, the aim of the study was to devise a method for predicting the morale of departments from objective data. The only description of the attitude measure is "A morale questionnaire was scored so that a quantitative score was available." There is no statement regarding the anonymity of the subjects. Three objective measures of efficiency were available. For each of 25 departments there were available three average measures of efficiency and one average morale score. When correlations based on group averages were computed they were found to range between .15 and .27. None of these is significant (our determination).

The Triple Audit studies (62) of the Industrial Relations Center at the University of Minnesota fall into this research design category. Here the firm was the basic unit of comparison. Since the number of firms studied was small, only seven, the authors advise that it is impossible to draw any conclusions about the relationships obtained.

The next series of studies reviewed here used the same design but differ from the three just mentioned in that they make some differentiation among employee attitudes. That is, they make some attempt to specify component parts.

The early work of Likert and Katz has been continued by them at the University of Michigan since the war. The prototype study was undertaken in 1947 in the Prudential Insurance Company and findings were reported in some detail in 1950 (34). Although the objectives were broad, an important portion of the study was devoted to exploring the relationships

between employee attitudes and productivity.

One and one-half hour free answer interviews covering 53 questions were held with 419 nonsupervisory clerical workers in 24 different sections. Responses were coded. The sections were arranged in parallel pairs in order to hold constant as many factors as possible. One set of 12 sections was designated as high productivity on the basis of production records while the parallel set was composed of low productivity sections. The authors note that productivity differences between the pairs were not great, rarely more than 10 per cent. Each of the high-low productivity pairs consisted of two sections handling the same type of work with the same type of people at the same job levels and were very similar on a number of factors.

A unique feature was the construction of four indices of attitudinal variables. The differentiations were made on a theoretical basis with some empirical confirmation for the relationship among the items used in each index. Four variables were specified: (a) pride in work group; (b) intrinsic job satisfaction; (c) company involvement; and (d) financial and job status satisfaction. Pride in work group was the most independent of the four; the remaining three inter-correlated around .4.

When these morale indices were related to productivity, only pride in work group showed a distinct relationship. Productivity groups were also differentiated by three specific attitude items not included in the morale indices.

A second study of similar design investigated these relationships among section hand employees of the Chesapeake & Ohio Railroad (35). Somewhat different morale items with more emphasis upon individual



items were used in intensive interviews. Productivity criteria consisted of over-all quality and quantity ratings by supervisors. There was some slight support for the previous finding of a relationship between pride in work group and productivity. The authors emphasize the lack of relationships found between employee attitudes and productivity.

Both Michigan studies, but particularly the insurance company one, may be studied with profit. The investigators are self-critical and also provide hypotheses for further investigation. The investigations are excellent examples of both the virtues and the shortcomings of survey techniques. Both illustrate attempts to measure a large number of variables with less precision perhaps than is ultimately desirable. These well-publicized studies are important in our present context because they have called into question a common assumption about an important relationship, have perhaps stimulated research elsewhere, and have produced a great amount of theorizing about motivation in industry.

Three recent investigations are patterned somewhat after the Michigan inquiries. Two studies by Comrey and associates (9, 10) are considerably less well reported. The findings, among Forest Service and Employment Service personnel, lend some slight support to the Michigan report of a relationship between attitudes toward the group and performance. Weschler (60) found a slightly negative relationship between a single-item index of job satisfaction and production among employees in two comparable groups of a Naval research laboratory. He obtained a similar result for a single item index of work group morale.

A recent study by Lawshe and Nagle (41) warrants extended com-

ment. Two hundred and eight non-supervisory office employees in 14 work groups at a plant of the International Harvester Company completed a 22-item questionnaire which was described by the authors as an attitude toward supervisor measure. The corrected split-half reliability was .92. There is no report regarding the anonymity of the respondents. The scores were related to group productivity.

A paired comparison rating of productivity by six plant executives was used. Each executive compared from 8 to 14 work groups under instructions to indicate "... The department in each pair which is, in your opinion, doing its job better." Ratings were converted to standard scores and averaged. The reliability of the means of all six raters was estimated to be .88. The authors are careful to point out that "one does not know for sure what the raters really had in mind when they rated." They suggest that "How little trouble the work group caused, whether or not it had the answers when called upon, whether or not it could cope with rush situations, and similar considerations are believed to have been the prime factors in the executives' ratings."

The average rating of each work group was correlated with average attitude toward the supervisor score in the work group. The resulting Pearson coefficient was .86, significant at the 1% level for  $N=14$ . This is, of course, a remarkable result. A whole superstructure of industrial psychology could well be erected on this finding; stranger things have happened. However, the authors sound a note of caution: "On the basis of this study it can be concluded only that the behavior of the supervisor, as perceived by the employees, is highly related to the productivity

of the group as perceived by higher management."

It occurs to us that it may be a misnomer to call the questionnaire an attitude questionnaire. It might well be considered to be a supervisor behavior- or performance-rating device. For example, the questions included such things as, does he: give you straight answers, avoid you when he knows you want to see him about a problem, criticize you for happenings over which you have no control, delay in taking care of your complaints, keep you informed, give you recognition, show interest in your ideas, follow through on his promises, explain to you the "why" of an error to prevent recurrence, give you sufficient explanation of why a work change is necessary. There is supporting evidence for a finding that supervisory performance is related to productivity (34). It might be suggested further that the obtained correlation really expresses the relationship between supervisor performance ratings by employees and supervisor performance ratings by the executives since the performance of a work group may be judged in part at least on the basis of the observation of the supervisor and certainly on his reports. We suggest that the finding is relevant to current work on supervisor performance and productivity but we are skeptical of its direct relationship to the area of research being examined in this review despite the title.

The results from the study design which we have described in this section are substantially in agreement with the previous findings of minimal or no relationship between employee attitudes and performance. They do supply the hint that morale, as a group phenomenon, may bear a positive relationship to performance on the job.

#### WITHDRAWAL FROM THE JOB

As indicated earlier we have differentiated between performance on the job and withdrawal from the job. In this section we briefly summarize the trend of the evidence when employee attitudes are related to some form of withdrawal from the job. Withdrawal is indicated by absence and tardiness, by accidents (under one assumption), and by turnover or employment stability.

##### *Absences*

*Individual analysis.* The individual analysis design has been used in four studies. In another of the Triple Audit studies, Yoder and associates (62) employed a 66-item employee attitude questionnaire which yielded a total score as an index of general attitude. Respondents apparently were unidentified. Absence and tardiness data were furnished by the respondents on the questionnaire face sheet. Five groups of employees were studied including office workers, department store personnel, and manufacturing employees. No statistically significant relationships were found between the attitude index and absences. One significant relationship was found for tardiness. Four others were insignificant.

In a study of worker attitudes toward merit rating, Van Zelst and Kerr (55) include data relevant to our topic. Three hundred and forty employees selected by their employers in 14 firms out of the 50 invited to participate furnished a self-report of their absences and tardinesses. Two Hoppock-type job satisfaction items were combined to give a single index. Respondents apparently were anonymous. Job satisfaction correlated .31 with a favorable absentee record and .26 with a favorable tardiness record. These are significant at the 1% level.

Bernberg (2) used four different measures of absence and tardiness which had split-half reliabilities in the seventies. These data were taken from the company records. The inter-correlations with the four measures of employee attitudes of the 890 aircraft plant workers ranged from  $-.05$  to  $+.07$ .

*Group analysis.* The group analysis design also has been used in four studies. Giese and Ruter (20) found an insignificant relation between tardiness and a single morale score when the group averages of employees in 25 departments of a mail order house were correlated. However, the correlation of  $-.47$  between the morale index and absences is significant at the 5% level (our determination).

Kerr and associates (38) obtained mean scores on his 10-item Tear Ballot job satisfaction blank for the employees in 30 departments of a Chicago plant presumably under conditions of anonymity. They used six measures of absenteeism. This is a major contribution of the study since there are numerous problems in indexing absenteeism. The importance of such an analytical approach is evident when it is observed that job satisfaction correlated  $.51$  with total absenteeism rate but correlated  $-.44$  with unexcused absenteeism. One other relationship was statistically significant.

In their wartime morale studies, Katz and Hyman (31) used six specific attitude items. These morale indices were positively related to absenteeism. The magnitude of the relationship is typified by responses to this item: 44% of the workers disliking their jobs were categorized as absentees as compared with 36% who liked their jobs.

Perhaps the most extensive investigation has been made by Metz-

ner and Mann (46) of the Michigan Survey Research Center. The data were collected in the Detroit Edison Company according to a design similar to the Prudential and Chesapeake & Ohio studies. Anonymous questionnaires provided the attitudinal data. White-collar and blue-collar men and white-collar women were the subjects. The most striking finding was that there was no relationship between absences and attitudes toward any aspect of the work situation for white-collar women. Among white-collar men, 10 out of 15 attitudinal measures showed significant relationships at the 10% level. Eight of these were significant at the 5% or 1% levels. However, when job level or grade was controlled to some extent by grouping into high- and low-skill levels, there was practically no relationship between attitudes and absences for the high-skill level jobs for the seven items it was possible to study. A fairly consistent relationship remained for the low-skill level jobs. Among the 18 items used with the 251 blue-collar men, nine were significant at the 10% level or better, six being at the 5% level or better. Incidentally, these are all percentage differences among various absence categories and the adjacent category differences are not particularly impressive although the differences between the extreme categories are appreciable.

#### ACCIDENTS

Hill and Trist (25), English investigators, have recently suggested that "Accidents (may) be considered as a means of withdrawal from the work situation through which the individual may take up the role of absentee in a way acceptable both to himself and to his employing organization." Accidents are considered to involve the "quality of the relation-

ship obtaining between employees and their place of work." In an empirical test of this hypothesis they found accident rates to be positively associated with other forms of absences and to be most strongly associated with the least sanctioned forms of absence. Their study does not, of course, bear directly on our immediate concern. We include it to indicate a possible linkage with absence data.

*Group analysis.* We have found two studies on the relationship between employee attitudes and accidents. Stagner and associates (53) used a group analysis design to study the job satisfaction of railroad employees. A total of 715 employees in 10 divisional groups, 2 accounting offices, and 12 shops were included. Fifteen specific items of an apparently anonymously administered questionnaire were given arbitrary weights based on the percentage of employees checking and were summed to give a single job satisfaction index. Mean satisfaction scores by groups were correlated with group accident rates. The obtained correlations are negative and small. Surprisingly, the authors conclude that "We thus feel considerable confidence in the conclusion that working in a group with a high accident rate will tend to make the individual worker anxious, and reduce his satisfaction with his job." However, the correlations are not statistically significant (our determination) and the causal sequence indicated in the quote is speculation.

The Triple Audit (62) studies also considered accidents. Although the authors have entered a general disclaimer as to the significance of their group design findings, the accident finding is intriguing. Employees in the three firms with fewer than average accidents had a mean attitude score of 133 while the employees in

the three firms with more than average accidents had a mean score of 143. From these limited data it appears that there is a tendency for the firms with higher accident rates to have more favorable attitude scores. The data are interesting but should not be given disproportionate emphasis.

#### EMPLOYMENT STABILITY

*Individual analysis.* Employment stability remains to be studied. In a study comparing indirect and direct methods of appraising employee attitudes, Weitz and Nuckols (59) provide relevant although peripheral data. This is an individual analysis design. Two attitude questionnaires, one composed of 18 indirect items and one consisting of 10 direct questions, were mailed to more than 1,200 insurance agents representing one company in the southern states. Forty-seven per cent submitted answers. The respondents were identified. Total scores for each of the questionnaires were then related to survival during a one-year period. The direct method correlated .20, significant at the 1% level; the indirect method correlated insignificantly with survival. There was some sample bias resulting from the fact that a disproportionately small number of men, who subsequently terminated, responded.

Kerr (37) correlated total Tear Ballot job satisfaction scores obtained individually but without identification from 98 miscellaneous wage earners with an index of *self-reported* past job tenure (number of years on labor market divided by number of employers). The result for an unweighted total satisfaction score was .25, significant at the 5% level. Thus there seems to be a slight positive relationship between attitude toward present employment and past em-

ployment stability among the members of a heterogeneous group making self-reports on their employment records.

Van Zelst and Kerr (55) correlated total score on two Hoppock-type items with *employee reports* of previous job tenure. The obtained correlation was .09.

Friesen (16) recently has attempted to measure employee attitudes using an incomplete sentences technique. He developed four scales comprising a total of 81 items. These he labeled Working Situation, Work, Self, Leisure. He studied women office workers from one company with *N*'s ranging from 38 to 70. The blanks were signed. Split-half reliability coefficients for the four scales ranged from .68 to .82. Intercorrelations among the scales ranged from .26 to .72 with Working Situation and Work being the two most highly correlated. Friesen attempted to validate the scales by obtaining modified "Guess Who" ratings from seven to nine fellow employees for each member of his sample. These ratings had uncorrected reliability coefficients ranging from .57 to .78. Their obtained correlations with the attitude scales were moderately high being .59, .67, .45, and .52, respectively. When the four attitude scales were related to a criterion of employment stability (two or more years with each employer versus less than two years with each employer) the biserial correlations were .43, .53, .37, and .22, respectively. All were significant at the 4% level or better. This is a *retrospective* measure of employment stability.

Two attitude items, chance to make decisions on the job, and a feeling they were making or had made an important contribution to the success of the company, were significantly related to turnover in a study

by Wickert (61). This study has the limitation that the employees who had left the company were interviewed after their departure.

*Group analysis.* The group analysis design has been used in three studies previously described. In the Giese and Ruter (20) investigation, morale scores correlated  $-.42$  with a per cent turnover criterion for 25 departments. This is significant at the 5% level (our determination). Kerr (38) found the relationship between total Tear Ballot score and turnover in 30 departments to be  $-.13$  which is not significant. The Triple Audit studies (62) found average monthly turnover to be unrelated to attitudes in seven companies.

With respect to withdrawal from the job, then, there is some evidence, mainly from the group design studies, of a significant but complex relationship between employee attitudes and absences. The investigations reviewed here also lend some support to the assumption that employee attitudes and employment stability are positively related. The data on accidents and attitudes are extremely limited, but they do not support any significant relationships.

In summary, it appears that there is little evidence in the available literature that employee attitudes of the type usually measured in morale surveys bear any simple—or, for that matter, appreciable—relationship to performance on the job. The data are suggestive mainly of a relationship between attitudes and two forms of withdrawal from the job. This tentative conclusion, contrary as it is to rather widely held beliefs, warrants an attempt to identify and evaluate some of the factors which may account for these results.

We have chosen to comment on



some methodological considerations, followed by a discussion of theoretical issues, and concluding with some possible implications for future research.

#### METHODOLOGICAL CONSIDERATIONS

The methodological questions which might be raised about any field study are legion (15, 27, 28, 36, 44). We shall comment briefly on some methodological limitations of the studies we have reviewed and on some fairly general problems of analysis and design of field study which are relevant to our topic.

##### *Limitations of the Current Literature*

It is difficult to know whether inadequacies in research reports reflect faults in research design, or whether space limitations in journals force authors to omit a considerable part of the methodological detail of their studies. This problem is apparent in the following discussion of three methodological areas: sampling, measurements, and the general procedure of the study.

**Sampling.** In most industrial studies there is sampling of both respondents and items. With regard to the sampling of respondents, reports frequently fail to state how respondents were selected, the possible selective biases, or the population which the sample is supposed to represent. It is possible that conflicting results from two studies may reflect differences in the characteristics of the populations sampled. This would be of considerable interest theoretically and practically, but it is unlikely to be detected unless the respective populations are described in some detail.

With regard to sampling of items, it is not uncommon for questionnaires administered to industrial workers to contain 150 or more items, and for research reports to mention re-

sults from only a few of these items. One is left in doubt as to whether the items reported upon were selected on the basis of some theoretical analysis, in which case the other items presumably were not theoretically relevant to the issue in question, or whether the items were selected on the basis of statistical significance and the theoretical orientation developed in retrospect. The latter procedure obviously capitalizes on chance, and the results should be subjected to further research before being accepted with any great confidence. Unless researchers inform the reader when they have operated on an *ad hoc* basis the tentative nature of the results will not be apparent.

**Criterion measures.** Research reports sometimes fail to describe the specific measurements that were used! In addition, there is extreme diversity in the kinds of employee attitudes that are measured, and in the questionnaires and interview schedules that are used to identify or measure them. This means that disparity in results between two or more studies may sometimes reflect differences in operational definitions, rather than differences between the populations being studied. We shall deal in greater detail with the measurement of attitudes in a subsequent section. Here we shall concentrate for the moment on criteria of job performance and of withdrawal from the job.

It is not our purpose to present a definitive report on problems involved in the selection and measurement of criteria. This area is well covered in current books in the field (19, ch. 3; 54, ch. 5). However, it is obviously impossible to assess the relation between employee attitudes and job performance without some measure of the latter. Therefore, it is appropriate to dwell briefly on the criterion problem.

Let us consider, first, criteria of job performance. On production-line jobs as well as in sales and many other white-collar jobs, the criterion which seems, at first glance, best to reflect adequacy of job performance is productivity. Yet the measurement of productivity is much less simple than it may seem initially.

To be valid, a comparison of output between two individuals must equate the conditions under which the individuals operate. One salesman's territory may be potentially more fruitful than another's and his higher sales may cause, rather than result from, his greater satisfaction with his job. Two machines may vary in their potential output, in their state of repair, and in many other ways. Frequently there are, in addition, external restraints on productivity. Output in a factory may be determined by the speed of the assembly line or the speed of the machine, by the amount of material provided to an individual by some feeder line, or by the quality of the material being processed. Variation in situational factors such as these will affect total productivity no matter what the level of individual job performance.

Furthermore, on many jobs, white-collar as well as production-line, it is impossible to get a measure of productivity because a certain amount of work is required during the day and no more is produced, because adequate records are not kept, because the product depends on group rather than individual performance, or for a variety of other reasons.

Where output data are questionable or unobtainable as a performance criterion, it is sometimes suggested that other objective criteria such as quality of production as measured by number of errors, amount of scrap, and the like, be used. However, many of the above

considerations such as machine variation, availability of records, and comparability of materials processed also apply to quality data.

Relevance, reliability, freedom from contamination, and practicality are criterion requisites which are not easily satisfied even when objective performance indices are available.

In the absence of direct counting measures of job performance, researchers are likely to obtain subjective evaluations, usually by superiors in the organization, of the performance levels of individuals or groups. Such ratings frequently were used in the investigations we have reviewed. The limitations and the precautions necessary in the use of ratings are well known and adequately documented (19, ch. 4; 22, ch. 11). Problems such as the selection of factors to be rated, the errors of human judgment-making, contamination, and reliability plague the investigator who is forced to rely upon ratings.

Criterion selection is a problem also in studies attempting to predict absenteeism and turnover. Sex differences, differences of position in the organization, and similar factors influence absences in such a way that differences in absence rate between two groups may reflect many things other than differences in employee attitudes. The same thing may be said of turnover. General employment and wage levels, selective service policies, and similar factors will all affect the rate of turnover and may mask the effect of employee attitudes on turnover.

The selection of criteria, then, involves a choice among a number of possible measurements all of which may be affected by situational factors over which the investigator has little if any control.

*Procedural problems.* A frequent drawback of reports of the type sum-

marized in this paper is a failure to report reliability data for the measurements, both of attitudes and of criteria of performance.

For the most part, the validity of the measures employed is unreported. With respect to the criterion measures this takes the form of failure to discuss the relevance of the particular criteria used. Usually employee attitude indices are assumed to have some form of face validity; empirical validation is seldom attempted. The reader is expected to assume that the questionnaires measured what they were intended to measure. In view of the history of measurement in psychology, however, experienced readers may hesitate to make this assumption.

We are indebted to S. Rains Wallace for the suggestion that a spurious factor may be present in studies which employ interviews or the extensive interpretation of questionnaires for determining morale in groups and relating it to the rated efficiency of the group.<sup>5</sup> It is not always clear whether the interviewers or interpreters had foreknowledge of the rated efficiency and thus might have contaminated their attitude measures.

Another procedural defect of some industrial studies is the use of self-reports or similar criterion data rather than independently obtained measures. Thus, in studies of absence and of turnover, respondents may be asked to tell how many times they have been absent in the last six months or how many jobs they have held in the last five years. Absence material, at least, could be collected in a more straightforward and probably more valid manner by looking at the individual's attendance record unless an attempt is being made to preserve anonymity.

<sup>5</sup> S. R. Wallace. Personal communication. October, 1954.

Sometimes, furthermore, the data collection is unsupervised, or is conducted incidentally to some other operation within the company. While there is no intrinsic objection to such projects, experience indicates that they frequently suffer from errors and extraneous factors and the reader should be warned of their tentative nature.

Finally, an important procedural issue concerns the practical problem of whether or not to identify the individual respondents. If we are to relate employee attitudes to measures or estimates of individual performance, it is necessary to be able to identify individual workmen. Frequently this means that the subject is required to sign the questionnaire or is interviewed individually. Although this is a crucial problem it has not been widely investigated or, sometimes, even considered. Thus it merits extended discussion.

We have located seven studies in industrial settings which bear upon the question of the influence of anonymity upon morale or job satisfaction (3, 14, 17, 24, 29, 45, 58). Two studies from the Survey Research Center, University of Michigan, compare responses made by employees to morale items when the items are contained in an anonymous questionnaire and when they are part of an interview. In Kahn's (29) study, comparisons were available for 65 items which were worded identically and were presented in the same context and sequence. In general, his findings support the conclusion that identification, at least via interview, produces differential responses. For example, the questionnaire items elicited more expressions of criticism and dissatisfaction and more extreme responses. A somewhat similar study by Metzner and Mann (45), although much more limited, gave comparable results. Both studies contain possible

defects in the matching of respondents as well as insufficient attention to interviewer differences.

A related study by Wedell and Smith (58) compared interviewer ratings of responses and coded ratings of the interview protocols with questionnaire responses of approximately 200 employees of a chemical company on three attitude questions. Interviewer ratings of responses differed significantly from questionnaire responses on the two questions dealing with attitude toward company and toward job. Coded ratings of the interview protocols differed significantly from questionnaire responses on the item dealing with attitude toward the company. In each instance of significant findings the interviews produced the most favorable attitude ratings as compared to the employees' actual questionnaire responses. The most interesting finding was the differences among interviewers. By and large, the more experienced and better trained interviewers showed the greatest discrepancies! The interview versus questionnaire methodological studies do not, of course, provide a crucial test of the anonymity factor since many other variables may be operating to confound the results (27).

In an abstract, Evans (14) reports a study of salaried employees which he says "provided for testing the hypothesis that employee anonymity should be preserved in management-conducted surveys." One major finding of his early analysis indicated that "employees are not particularly concerned about anonymity." He concludes that "the unwarranted popular assumption that employees must be anonymous is open to serious question." The brevity of the report makes it difficult to evaluate the significance of this conclusion.

In an early investigation of the

problem in an industrial setting, Brayfield (3) compared the distribution of scores on a job satisfaction questionnaire obtained from female stenographers employed in private industry with those for a similar group employed in state civil service. The members of the industrial group identified themselves under good rapport conditions and the civil service group was anonymous. It was apparent by inspection that the two groups gave similar responses. The results are inconclusive particularly since the groups were not working under strictly comparable conditions.

Gadel and Kriedt (17) report briefly upon a test of the effect of anonymity upon employee attitude questionnaires. They found that the distributions of answers obtained under the differing conditions of anonymity and identification were almost identical for several groups of employees. The small differences present were inconsistent from group to group and from item to item.

In a carefully controlled study, Hamel and Reif (24) administered a 65-item employee attitude questionnaire to two random samples of all employees in a large department store. The members of one sample remained anonymous; those in the second sample identified themselves by name and department number. The mean scores for the two groups were not significantly different. When the group responses to each individual item were compared only two item responses in 65 were significantly different. For the number of comparisons involved the authors state that "this finding is slightly smaller than expected."

The results from studies outside the industrial situation using such diverse materials as attitude scales, opinion polls and ballots, and personality tests are equivocal. Hyman

(27, p. 185), in a considered treatment of the issue, says "the foregoing discussion should serve to make clear the complexity in estimating the nature and direction of effects due to identification or anonymity of the respondent." Our own review of the dozen or so relevant nonindustrial investigations along with the industrial studies already cited leads us to agree with Kahn's (29, p. 8) conclusion that "the studies on the effect of anonymity of response conducted during the past two decades compel us to conclude that there is no predictable effect of anonymity *per se*."

We would sound the caution that, in any study of the problem, it is necessary to differentiate literal anonymity from psychological anonymity, as Hyman (27) has suggested, since even the questionnaire studies usually require identification at least to the extent of naming one's department, and sometimes even more extensive data are called for. These requirements and situational factors may influence responses to what the investigator considers to be anonymously answered materials.

We do hazard the opinion, however, that in a situation where rapport has been carefully nurtured in all steps of the investigation, the identification of *questionnaire* materials will not necessarily result in serious distortion of responses. Yet the dilemma remains. The investigator who asks for identification on the questionnaire, and perhaps especially the investigator who relies upon the interview, risks distortion in responses, while the one who does not ask for identification is unable to make certain crucial analyses. The ethical and practical consequences of coding and disguised identification of respondents are such as to render that particular solution debatable.

#### *General Problems of Analysis and Design*

We shall be concerned here with a discussion of scaling and statistical techniques commonly in use in psychology, with the general problem of operational definitions as applied to the specific area of industrial research, and with the validity of certain group measures.

*Statistical and scaling techniques.* In the typical industrial study the psychologist asks a group of workers to fill out a questionnaire which is designed to measure "morale," he ranks the same employees on some aspects of their job performance, and he then runs a *t* test or a product-moment correlation between morale, on the one hand, and performance on the other. We believe that too little recognition has been given to the assumptions about the nature of the data which are built into these measurements and into the statistical tests.

Thus, for example, the typical morale scale makes the assumption that a unidimensional continuum is being measured and that the items of the scale are equally important in contributing to the individual's position on the continuum. Except for item analysis in the initial construction of the scales, it is doubtful whether the validity of these assumptions is ever tested and item analysis does not guarantee unidimensionality, nor is it often used for differential weighting of items. The use of mean scores and of product-moment correlation adds such further assumptions as that the intervals between points on the scale are of equal magnitude, that the parent population is normally distributed on the variables concerned, that homoscedasticity exists in the case of the Pearsonian coefficient, and so on. To the extent that these assumptions are violated,



the results become ambiguous. It is doubtful if the validity of the assumptions is routinely tested.

However, since these comments hold true generally for much research in all fields of psychology, it would be unreasonable to expect a different type of analysis among observers studying morale.

*Measurement of morale.* McNemar (44) has commented at length upon the tendency for attitude researchers to "measure" a concept by applying a particular label (say conservatism-radicalism) to a set of questions without concern for other operational definitions of the same concept. "Morale" has probably been as greatly subject to this procedure as any other concept in psychology today. The lack of consistent definition of the term "morale" has led us to title this article "Employee attitudes and employee performance" in the hope that the term "attitude" provides the general and ambiguous connotation that is required to describe measurement in this field.

One likely cause of this proliferation of operational definitions is that morale is a global concept. Researchers tend to consider any area of satisfaction as important for the employee's over-all morale, for his satisfaction with his job and his company. This has led to attempts, logically and through factor analysis, to specify particular aspects of morale. Frequently subareas of morale are identified, such as satisfaction with the work group, satisfaction with pay, with supervision, with promotion, with type of work, with physical aspects of the job, or with any other component of the working environment. Questions are then developed to provide an indication of the degree of employee satisfaction in each of these areas.

This move toward increased speci-

fication of the components of the concept, and toward the development of unidimensional scales probably should be applauded. However, a problem arises of how to use the subscales. Frequently the items within a subscale are summed to provide a score on the scale, and then the individuals' positions on the subscales are summed to provide an "index" of over-all morale. In such an event, the subscales constitute simply a means of weighting items, and subscales are assumed to be equally important in producing whatever over-all morale is supposed to be. The possibility remains that a more fruitful method of analysis would be to consider the subscales independently or as configurations rather than to combine them additively.

The reason usually given for using multi-item scales in measuring morale is the same as that given for using many items in tests of ability: the greater reliability of longer scales. It is assumed that increasing the number of attitude items increases the likelihood that respondents will be ordered in the same way on a second administration of the questionnaire. Yet the analogy to ability tests is not necessarily a good one. While there may be an infinite population of arithmetic problems, for example, from which test items may be chosen, the inclusion of additional items in attitude scales is likely to introduce new attitude dimensions. Unless these dimensions are highly correlated, the scores on the questionnaire may be ambiguous. For example, a person could achieve a moderate morale score by indicating moderate satisfaction with his pay and with the type of work he is doing, or by indicating great satisfaction with the type of work he is doing and great dissatisfaction with his pay. It is not unlikely that these different patterns

of response would be differently related to job performance, to absence, and to turnover.

Concern with unidimensionality, rather than with reliability as traditionally conceived, would lead to the use of such scaling techniques as Guttman's, Loewinger's, or Coombs's, or to the use of single items. It is interesting that many of the studies we have reported above have made at least partial use of the single-item approach (32, 35, 46, 48, 60, 61).

*Use of group measurements.* Frequently, members of a work group or a department fill out morale questionnaires individually, but have group, rather than individual, performance or productivity records. In such situations it is customary to test the significance of the difference between the morale scores of the two groups to determine whether the higher producing group also has higher morale. We shall comment upon only one aspect of this procedure by pointing out that a relationship which exists at the individual level between satisfaction and productivity may be obscured when the individuals are lumped together. There may be a positive relationship between satisfaction and performance within each group even though the two groups as a whole do not differ significantly with regard to mean satisfaction. Such a relationship obviously is not revealed in this type of analysis and lack of individual performance records makes the appropriate analysis impossible.

The degree to which these and other methodological issues may have beclouded the relationships between employee attitudes, performance, absenteeism, and employment stability is an open question. Certainly, as must be obvious to the careful reader, the studies we have reviewed are subject to some or, occasionally, most of

the shortcomings described above. However, the scarcity of relationships, either positive or negative, demonstrated to date even among the best designed of the available studies leads us to question whether or not methodological changes alone would lead to a substantial increase in the magnitude of the obtained relationships. We are led, then, from consideration of the current status of research on this topic to a discussion of the relationships on the conceptual level. Much of what we will say has previously been elaborated by Katz and Kahn (33) and by Morse (47).

#### THEORETICAL CONSIDERATIONS

##### *Morale as an Explanatory Concept in Industrial Psychology*

One principal generalization suffices to set up an expectation that morale should be related to absenteeism and turnover, namely, that organisms tend to avoid those situations which are punishing and to seek out situations that are rewarding. To the extent that worker dissatisfaction indicates that the individual is in a punishing situation, we should expect dissatisfied workers to be absent more often and to quit the job at a higher rate than individuals who are satisfied with their work. Since the general proposition about the effects of reward has received a great amount of verification in psychology, it is not strange that it has been carried to the analysis of absenteeism and turnover.

A plausible connection between satisfaction and performance on the job is less obvious. Let us consider specifically the possible relationship between satisfaction and productivity. Under conditions of marked dissatisfaction it is likely that low productivity may serve as a form of aggression which reflects worker hostility toward management. But the hypothesis that production should

increase monotonically with increases in satisfaction apparently rests on the assumption that the worker will demonstrate his gratitude by increased output, or that the increased satisfaction frees certain creative energies in the worker, or that the satisfied employee accepts management's goals, which include high production.

In any event, it is commonly hypothesized that, whatever the causes, increased satisfaction makes workers more motivated to produce. Given this condition, it should follow that increased productivity can be attained by increasing worker satisfaction. We are going to advance the proposition that the motivational structure of industrial workers is not so simple as is implied in this formula. We feel that research workers have erred by overlooking individual differences in motivations and perceptions because of their concern with discovering important and applicable generalizations. Most of what follows is an effort to point out areas in which differences between workmen may make a difference in their adjustment to the situation.

At the outset let us make it clear that we expect the relation between satisfaction and job performance to be one of concomitant variation, rather than cause and effect. It makes sense to us to assume that individuals are motivated to achieve certain environmental goals and that the achievement of these goals results in satisfaction. Productivity is seldom a goal in itself but is more commonly a means to goal attainment. Therefore, as G. M. Mahoney has suggested,<sup>6</sup> we might expect high satisfaction and high productivity to occur together when productivity is perceived as a path to certain im-

portant goals and when these goals are achieved. Under other conditions, satisfaction and productivity might be unrelated or even negatively related.

In the light of this consideration, we shall center our discussion on an analysis of industrial motivation as it relates specifically to employee satisfaction and to productivity.

For the sake of convenience we may distinguish between threats and rewards as incentives to productivity. Goode and Fowler (21) have described a factory in which morale and productivity were negatively related but productivity was kept high by the continuance of threats to workers. Here the essential workers—people with considerable skill—were marginal to the labor force because of their sex or because of physical handicaps. Since the plant was not unionized, it was possible for management to demand high productivity from these workers on threat of discharge. This meant that the workers, although most dissatisfied with their jobs, produced at a very high rate because of the difficulty they would face in finding another position should they be discharged.

There is little doubt that threat was widely used as a motivating device in our own society in the past and is presently used in more authoritarian societies. However, it is doubtful if any great amount of at least explicit threat is currently used by industries in this country in efforts to increase productivity or reduce absenteeism. First of all, considerable change has occurred in management philosophy over the past fifty years, and such tactics are repugnant to many industrial concerns. Secondly, the growth of unions has virtually outlawed such tendencies except in small, semi-marginal in-

<sup>6</sup> G. M. Mahoney. Personal communication. March, 1953.

dustries which are not unionized.

Threats of discharge, then, probably do not operate as incentives unless the worker falls considerably below the mean in quantity and/or quality of output. For a number of reasons management has settled upon rewards for motivating workers to produce, including such tangible incentives as increased pay and promotion, as well as verbal and other symbolic recognition. Let us examine whether this system of rewards actually provides motivation for increased productivity by the worker.

It is a commonplace observation that motivation is not a simple concept. It is a problem which may be attacked at a number of different levels and from many theoretical points of view. Whatever their theoretical predilection, however, psychologists generally are agreed that human motivation is seldom directed only toward goals of physical well-being. Once a certain minimum level of living has been achieved, human behavior is directed largely toward some social goal or goals. Thus, in our own society, goals such as achievement, acceptance by others, dominance over others, and so on, probably are of as great concern to the average workman as the goals of finding sufficient food and shelter to keep body and psyche together.

We assume that social motives are of considerable importance in industry. We assume, further, that the goals an individual pursues will vary, depending upon the social systems within which he is behaving from time to time. Most industrial workers probably operate in a number of social systems. Katz and Kahn (33) suggest four such systems: first, the system of relations outside the plant; and, within the plant, the systems of relationship with fellow workers on the job, with members of

the union, and with others in the company structure. We may ask whether job performance, and particularly productivity, is a path to goal achievement within these various sets of social relations.

*Outside the plant.* It is often argued that any worker who is motivated to increase his status in the outside community should be motivated toward higher productivity within the plant. Productivity frequently leads directly to more money on the job, or involves movement to jobs with higher prestige or with authority over others. If productivity does result in such in-plant mobility, increased output may enable the individual to achieve a higher level of living, to increase his general status in the community, and to attempt such social mobility as he may desire. In this way productivity may serve as a path to the achievement of goals outside the plant.

The operation of this chain of relationships, however, depends not only upon the rewards given the high producer, but also upon the original motivation of the workman to increase his status position in the outside community. The amount of status motivation among production-line employees is open to question. Certainly the findings of Warner (57), Davis and Gardner (12), and others (6, 11, 13), indicate that there are systematic differences in the goals which are pursued in the different segments of our society. It is not impossible that a very large proportion of America's work force is only minimally motivated toward individual social achievement. The assumption that such a motivation does exist may reflect in considerable part a projection of certain middle-class aspirations onto working-class employees.

Furthermore, it is not unlikely

that the reference group against which an individual workman evaluates his success may be only a segment of the community, rather than the community as a whole. An individual whose accomplishments are modest at best when compared with the range of possible accomplishments in the community may have a feeling of great accomplishment when he compares his achievements with those of others in his environment. If this is true, and if he desires to continue to operate within this segment of society, any further increase in rewards within the plant might lead to his exclusion from personally important groups outside the plant rather than to increased prestige in such groups.

Finally, there are many goals outside the industrial plant which may be socially rewarding to the individual and which require only minimal financial and occupational rewards inside the plant. Active participation in veterans' organizations, in churches, in recreational programs and similar activities may be and frequently are carried out by individuals at all positions in the industrial hierarchy. As a matter of fact, to the extent that the individual receives extensive social rewards from such activities he may have only slight interest in his work on the job, and he may continue to remain in industry only to maintain some minimum economic position while carrying out his outside functions. For such an individual, high productivity may lead to no important goals.

*Relations with other workers in the plant.* The studies by Elton Mayo and his associates (43, 50, 51) introduced the work group into the analysis of industry, and a wealth of subsequent investigations have confirmed the importance of on-the-job groups. Throughout these studies

has run the observation that members of the work group develop group standards of productivity and attempt to force these standards upon those workmen who deviate. Thus, in the Bank Wiring Room (51) it was the socially maladjusted individual, the deviant from the work group, who maintained a level of production above that of the group even though his native ability was considerably below that of many of the others.

Mathewson's (42) classic study of restriction of output among unorganized workers was an early demonstration of the operation of group norms.

Schachter and associates (52) have conducted an experiment which indicates that in cohesive groups an individual's productivity may be either raised or lowered, depending upon the kind of communications directed toward him by congenial co-workers. In an actual factory setting, Coch and French (8) presented existent groups with evidence that a change in job methods and in productivity was necessary if the factory was to remain in a favorable position relative to other, competing factories. These groups, through group discussion, arrived at a decision as to the proper job set up, and modified the group judgment of "fair" output markedly upward.

There is evidence, then, that the level of performance on the job frequently depends upon a group norm, and that performance level may be changed by changing the group norm in a direction desired by management. This change in the norm probably results from a conviction among the workers that higher production is in their own interest as well as management's, i.e., that their interests and management's interests coincide. This raises the perplexing question



of whether, with regard to productivity, the interests of management and labor do, in fact, coincide.

Management, presumably, is interested in higher production as a way of reducing the ratio of cost to output, and thereby bettering management's financial and competitive position. In an expanding market, the argument goes, this makes possible the expansion of the company, increased wages, a larger labor force, and general prosperity not only for the corporation but for the employees as well.

The case may not be so attractive to the workers, especially when the market is not expanding and demand for the product is constant, nearly constant, or declining. In this event, higher productivity per worker means that fewer people are required for the same level of output, or that fewer hours are worked by the same number of workers. In either case, many workers may lose, rather than gain, by the increase in productivity. It may be argued that in normal times such individuals usually find fairly rapid employment in some other segment of the economy. However true this may be, from the viewpoint of the individual workman this involves a considerable disruption in working habits and in his social life in general, and is to be avoided wherever possible. Viewed in this light the interests of management and labor are inimical.

As psychologists we steer clear of such arguments. But we should be sensitive to the fact that the question is a debatable one, that a final decision will probably rest upon values rather than data, that each side is capable of convincing arguments, and that the perception of a certain inevitable conflict of interests between the two groups is honestly and intelligently held by many people. We should also recognize that any

reduction in work force after a joint labor-management effort to increase productivity will likely be interpreted as resulting from the increased productivity, and may lead to a future avoidance not only of high productivity levels but also of labor-management cooperation.

At any rate, we often find that individual workers interpret higher productivity as counter to the interests of the employees. To the extent that this perception constitutes a group norm, such motives as are rewarded through the individual's social relationships with other workmen may be blocked by increased productivity. In such cases, productivity may serve as a path to certain goals, but as a block to social acceptance.

*The union structure.* One system of relationships of considerable importance in many industrial concerns is the union. In many companies much of what was said in the preceding section may be extended to refer also to the relations of the worker in the system of social relations within the union.

In some plants high productivity is not a deterrent to active union participation. Nevertheless, it probably is true that productivity is seldom a prerequisite for advancement within the union hierarchy. If the individual is oriented toward the union structure, it is unlikely that high productivity will serve as a path to such goals, whatever its effect on other goals he may pursue.

*The company structure.* We have indicated above that many of the worker's social motives outside the plant, as well as his desires for in-plant associations with fellow workmen and within the union, may be only slightly affected by increases in productivity and sometimes may be blocked by increased productivity.

The apparent range of goals that a worker may have is so wide that productivity may be a path to only a few of them.

However, workers are often motivated toward goals within the plant such as turning out a quality product, higher wages, and promotion. Let us examine the relationship between satisfaction and productivity for workers who are motivated toward these in-plant goals.

At the start it is evident that productivity and quality are sometimes mutually exclusive. If the individual must concentrate on maintaining high quality work, speed of production probably plays a secondary role. Conversely, if he must emphasize speed, quality often must be reduced to some degree. The speed-quality dilemma is sometimes resolved by making the individual work units so routine and concerned with such minute changes in the material that increased speed will not affect the quality of the product. However, if a worker is more highly motivated when he is performing some meaningful job, the above procedure may be resolving one dilemma by raising another. At any rate, the artisan, motivated toward the goal of quality, may be highly satisfied with his job while turning out a very limited number of finished pieces per unit of time. If he is forced to increase productivity and lower in some measure the quality, we might expect his satisfaction to decrease. For such a person satisfaction and productivity would be negatively related.

Consider now the individual who is motivated toward higher wages and promotion. While these rewards may not be exclusively dependent upon job performance, at the same time productivity and other aspects of performance often are weighted heavily at promotion time in most

companies. In other words, productivity and other aspects of job performance constitute a path to the goal of promotion and wage increases.

Now it is likely that people with aspirations to change position in the company structure will often be quite dissatisfied with their present position in the company. Aspiration to move within a system implies not only a desire for some different position in the future, but some degree of dissatisfaction with the position one is presently occupying. The amount of dissatisfaction probably depends upon the length of time the individual has occupied this position. Thus, although productivity may be a path to the goal, failure to achieve the goal to date may result in dissatisfaction and the high producer may be less satisfied than the low producer.

Evidence sustaining this point of view is to be found in Katz and associates' (34) report of a large insurance company in which the best, most productive workers were also considerably more critical of company policy than were less productive workers. S. Lieberman reports a similar finding in a large appliance factory.<sup>1</sup> A year after all workers in the factory had filled out a questionnaire, Lieberman compared the earlier responses of those who had been promoted to foreman with a matched group of workers who were not promoted. Those promoted had been significantly less satisfied with company practices at the earlier time than had the control group.

Once again the question arises as to what is meant by satisfaction. It may be that extremely high satisfaction is indicative of a certain amount of complacency, a satisfaction with the job as it is, which may be only

<sup>1</sup> S. Lieberman. Personal communication. July 15, 1954.

slightly related to job performance, if it is related at all. On the other hand, individuals who are highly motivated may perceive productivity as a path to their goals, but may also be more realistically critical of whatever deficiencies exist within the organization. They may feel, in addition, that their output is not being rewarded as rapidly as it deserves.

#### *Implications for Future Research*

We have arrived at two conclusions: first, that satisfaction with one's position in a network of relationships need not imply strong motivation to outstanding performance within that system, and, second, that productivity may be only peripherally related to many of the goals toward which the industrial worker is striving. We do not mean to imply that researchers should have known all along that their results would be positive only infrequently and in particular circumstances. We have been operating on the basis of hindsight and have attempted to spell out some of the factors which may have accounted for the failure of industrial investigators to find positive relationships in their data.

However, certain implications seem logical from the foregoing sections of this report. Foremost among these implications is the conclusion that it is time to question the strategic and ethical merits of selling to industrial concerns an assumed relationship between employee attitudes and employee performance. In the absence of more convincing evidence than is now at hand with regard to the beneficial effects on job performance of high morale, we are led to the conclusion that we might better forego publicizing these alleged effects.

The emphasis on predicting job performance, and particularly productivity, rests upon the acceptance

of certain values. That is, the many studies that have used productivity as the criterion to be predicted have been performed because productivity has direct economic value to industry, and, presumably, to society at large. But the fact that it has economic value does not mean that job performance is the only, or even the most important, aspect of organizational behavior. From the viewpoint of studying, analyzing, and understanding the industrial setting and individual reactions thereto, productivity and other aspects of job performance may be only one of several important factors. It would seem worthwhile to study the causes, correlates, and consequence of satisfaction, per se. It seems possible, for example, that conditions conducive to job satisfaction will have an effect on the quality of the workman drawn into the industry, the quality of job performance, and the harmony of labor-management relations. Such potential correlates, among others, merit exploration.

Another potentially fruitful approach involves studying the differential effect of particular kinds of management practices upon the attitudes and performances of workers with different motives, aspirations, and expectations. The appropriate questions may concern how, for particular workers, productivity comes to be perceived as instrumental to the achievement of some goals but not others, while for other workers a different perception develops.

The experimental approach has largely been neglected in this area of industrial research, yet the control of variables that it provides seems essential to the development and refinement of our knowledge in the area. Certainly, where experimentation has been used, as by Schachter and associates (52) and by Coch and

French (8), the results have been both enlightening for the understanding of present problems and encouraging for its future application. As our concepts become increasingly precise, we may expect an increased use of experimentation both within the industrial setting and in the laboratory.

Perhaps the most significant conclusion to be drawn from this survey of the literature is that the industrial situation is a complex one. We have suggested that an analysis of the situation involves analysis not only of the individual's relation to the social system of the factory, the

work group, and the union, but the community at large as well. It is important to know what motives exist among industrial workers, how they are reflected in the behavior of the workers, and how the motives develop and are modified within the framework of patterned social relationships in the plant and in the larger community.

We seem to have arrived at the position where the social scientist in the industrial setting must concern himself with a full-scale analysis of that situation. Pursuit of this goal should provide us with considerable intrinsic job satisfaction.

#### REFERENCES

1. BAXTER, B., TAAFFE, A. A., & HUGHES, J. F. A training evaluation study. *Personnel Psychol.*, 1953, 6, 403-417.
2. BERNBERG, R. E. Socio-psychological factors in industrial morale: I. The prediction of specific indicators. *J. soc. Psychol.*, 1952, 36, 73-82.
3. BRAYFIELD, A. H. *The interrelation of measures of ability, aptitude, interests, and job satisfaction among clerical employees*. Unpublished doctor's dissertation, Univer. of Minnesota, 1946.
4. BRAYFIELD, A. H., & ROTHE, H. F. An index of job satisfaction. *J. appl. Psychol.*, 1951, 35, 307-311.
5. BRODY, MILDRED. *The relationship between efficiency and job satisfaction*. Unpublished master's thesis, New York Univer., 1945.
6. CENTERS, R. *The psychology of social classes*. Princeton: Princeton Univer. Press, 1949.
7. CHASE, F. S. Factors for satisfaction in teaching. *Phi Delta Kappan*, 1951, 33, 127-132.
8. COCH, L., & FRENCH, J. R., JR. Overcoming resistance to change. *Hum. Relat.*, 1948, 1, 512-532.
9. COMREY, A. L., PFIFFNER, J. M., & BEEM, HELEN P. Factors influencing organizational effectiveness. I. The U. S. Forest Survey. *Personnel Psychol.*, 1952, 5, 307-328.
10. COMREY, A. L., PFIFFNER, J. M., & BEEM, HELEN P. Factors influencing organizational effectiveness. II. The
- Department of Employment Survey. *Personnel Psychol.*, 1953, 6, 65-79.
11. DAVIS, A. *Social class influences upon learning*. Cambridge: Harvard Univer. Press, 1948.
12. DAVIS, A., GARDNER, B. B., & GARDNER, MARY R. *Deep south: A social and anthropological study of caste and class*. Chicago: Univer. of Chicago Press, 1941.
13. ERICSON, MARTHA C. Social status and child rearing practices. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947. Pp. 494-501.
14. EVANS, C. E. Item structure variation as a methodological problem in an employee survey. *Amer. Psychologist*, 1949, 4, 280. (Abstract)
15. FESTINGER, L., & KATZ, D. (Eds.) *Research methods in the behavioral sciences*. New York: Dryden Press, 1953.
16. FRIESEN, E. P. The incomplete sentences technique as a measure of employee attitudes. *Personnel Psychol.*, 1952, 5, 329-345.
17. GADEL, MARGUERITE S., & KRIEDT, P. H. Relationships of aptitude, interest, performance, and job satisfaction of IBM operators. *Personnel Psychol.*, 1952, 5, 207-212.
18. GARRETT, H. E. *Statistics in psychology and education*. New York: Longmans, Green, 1947.
19. GHISELLI, E. E., & BROWN, C. W. *Personnel and industrial psychology*. New York: McGraw-Hill, 1948.

20. GIESE, W. J., & RUTER, H. W. An objective analysis of morale. *J. appl. Psychol.*, 1949, **33**, 421-427.
21. GOODE, W. J., & FOWLER, I. Incentive factors in a low morale plant. *Amer. sociol. Rev.*, 1949, **14**, 618-624.
22. GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
23. HABBE, S. Job attitudes of life insurance agents. *J. appl. Psychol.*, 1947, **31**, 111-128.
24. HAMEL, L., & REIF, H. G. Should attitude questionnaires be signed? *Personnel Psychol.*, 1952, **5**, 87-91.
25. HILL, J. M. M., & TRIST, E. L. A consideration of industrial accidents as a means of withdrawal from the work situation. *Hum. Relat.*, 1953, **6**, 357-380.
26. HOUSER, J. D. *What the employer thinks*. Cambridge: Harvard Univer. Press, 1927.
27. HYMAN, H. H. *Interviewing in social research*. Chicago: Univer. of Chicago Press, 1954.
28. JAHODA, MARIE, DEUTSCH, M., & COOK, S. W. (Eds.) *Research methods in social relations*. New York: Dryden Press, 1951.
29. KAHN, R. L. *A comparison of two methods of collecting data for social research: The fixed-alternative questionnaire and the open-ended interview*. Unpublished doctor's dissertation, Univer. of Michigan, 1952.
30. KATZ, D. Morale and motivation in industry. In W. Dennis (Ed.), *Current trends in industrial psychology*. Pittsburgh: Univer. of Pittsburgh Press, 1949. Pp. 145-171.
31. KATZ, D., & HYMAN, H. Industrial morale and public opinion methods. *Int. J. Opin. Attit. Res.*, 1947, **1**, 13-30.
32. KATZ, D., & HYMAN, H. Morale in war industry. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947. Pp. 437-447.
33. KATZ, D., & KAHN, R. L. Some recent findings in human relations research in industry. In G. E. Swanson, T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1952. Pp. 650-665.
34. KATZ, D., MACCOBY, N., & MORSE, NANCY. *Productivity, supervision and morale in an office situation*. Univer. of Michigan: Survey Research Center, 1950.
35. KATZ, D., MACCOBY, N., GURIN, G., & FLOOR, L. G. *Productivity, supervision and morale among railroad workers*. Univer. of Michigan: Survey Research Center, 1951.
36. KENDALL, PATRICIA, L., & LAZARSFELD, P. F. Problems in survey analysis. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in social research*. Glencoe, Ill.: Free Press, 1950.
37. KERR, W. A. On the validity and reliability of the job satisfaction Tear Ballot. *J. appl. Psychol.*, 1948, **32**, 275-281.
38. KERR, W. A. Summary of validity studies of the Tear Ballot. *Personnel Psychol.*, 1952, **5**, 105-113.
39. KORNHAUSER, A., & SHARP, A. Employee attitudes: suggestions from a study in a factory. *Personnel J.*, 1932, **10**, 393-401.
40. KRISTY, N. F. *Criteria of occupational success among post office counter clerks*. Unpublished doctor's thesis, Univer. of London, 1952.
41. LAWSHE, C. H., & NAGLE, B. F. Productivity and attitude toward supervisor. *J. appl. Psychol.*, 1953, **37**, 159-162.
42. MATHEWSON, S. B. *Restriction of output among unorganized workers*. New York: Viking Press, 1931.
43. MAYO, E. *The social problems of an industrial civilization*. Cambridge: Graduate School of Business Administration, Harvard Univer., 1945.
44. MCNEMAR, Q. Opinion-attitude methodology. *Psychol. Bull.*, 1946, **43**, 289-374.
45. METZNER, HELEN, & MANN, F. A limited comparison of two methods of data collection: the fixed alternative questionnaire and the open-ended interview. *Amer. sociol. Rev.*, 1952, **17**, 486-491.
46. METZNER, HELEN, & MANN, F. Employee attitudes and absences. *Personnel Psychol.*, 1953, **6**, 467-485.
47. MORSE, NANCY C. *Satisfactions in the white-collar job*. Univer. of Michigan: Survey Research Center, 1953.
48. MOSSIN, A. C. *Selling performance and contentment in relation to school background*. New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1949.
49. RAUBE, S. A. *Experience with employee attitude surveys* (Studies In Personnel Policy, No. 115). New York: National Industrial Conference Board, 1951.
50. ROETHLISBERGER, F. J. *Management and morale*. Cambridge: Harvard Univer. Press, 1943.
51. ROETHLISBERGER, F. J., & DICKSON, W. J. *Management and the worker*. Cam-



- bridge: Harvard Univer. Press, 1939.
52. SCHACHTER, S., ELLERTSON, N., MCBRIDE, D., & GREGORY, D. An experimental study of cohesiveness and productivity. *Hum. Relat.*, 1951, 4, 229-238.
53. STAGNER, R., FLEBBE, D. R., & WOOD, E. F. Working on the railroad: a study of job satisfaction. *Personnel Psychol.*, 1952, 5, 293-306.
54. THORNDIKE, R. L. *Personnel selection*. New York: Wiley, 1949.
55. VAN ZELST, R. H., & KERR, W. A. Workers' attitudes toward merit rating. *Personnel Psychol.*, 1953, 6, 159-172.
56. VITELES, M. S. *Motivation and morale in industry*. New York: Norton, 1953.
57. WARNER, W. L., & LUNT, P. S. *The social life of a modern community*. New Haven: Yale Univer. Press, 1941.
58. WEDELL, C., & SMITH, K. U. Consistency of interview methods in appraisal of attitudes. *J. appl. Psychol.*, 1951, 35, 392-396.
59. WEITZ, J., & NUCKOLS, R. C. The validity of direct and indirect questions in measuring job satisfaction. *Personnel Psychol.*, 1953, 5, 487-494.
60. WESCHLER, I. R., KAHANE, M., & TANNENBAUM, R. Job satisfaction, productivity and morale: a case study. *Occupational Psychol.*, 1952, 26, 1-14.
61. WICKERT, F. R. Turnover, and employees' feelings of ego-involvement in the day-to-day operations of a company. *Personnel Psychol.*, 1951, 4, 185-197.
62. YODER, D., HENEMAN, H., JR., & CHEIT, E. F. *Triple audit of industrial relations*. Minneapolis: Univer. of Minnesota Press, 1951.

Received July 27, 1954.

# THE "POST-MORTEM" TESTING OF EXPERIMENTAL COMPARISONS

RICHARD B. McHUGH  
*Iowa State College*

AND

DOUGLAS S. ELLIS  
*Air Force Personnel and Training  
Research Center, Denver*

The advantage of using an efficient experimental design which permits an over-all analysis of variance  $F$  test has been frequently pointed out, for example, in the recent texts of Johnson (3), Lindquist (5), and Walker and Lev (8). As illustration, a portion of the reading experiment data

ing, he frequently wishes to pursue the corresponding effects further by contrasting means or totals from certain groups of treatments. For example (operating at the 1% level), since  $F$  for *remedial instruction methods* is significant,<sup>2</sup> then the psychologist might wish to know if the significance

TABLE 1  
ANALYSIS OF VARIANCE OF READING IMPROVEMENT SCORES FROM  
BURT AND LEWIS DATA (1)

Source	df	SS	MS	F	F <sub>1%</sub>
Methods of Remedial Instruction	3	880.1	293.4	9.31	4.46
Methods of Original Teaching	3	230.6	76.9	2.44	4.46
Interaction	9	764.9	85.0	2.70	3.01
Error	32	1007.6	31.5		

TABLE 2  
TOTALS AND MEANS FOR THE FOUR METHODS OF REMEDIAL INSTRUCTION

Methods of Remedial Instruction	Total	Mean	Number of Cases
Alphabetic	$T_1 = 1267.3$	$M_1 = 105.61$	$N_1 = 12$
Kinesthetic	$T_2 = 1347.2$	$M_2 = 112.27$	$N_2 = 12$
Phonic	$T_3 = 1247.2$	$M_3 = 103.93$	$N_3 = 12$
Visual	$T_4 = 1368.6$	$M_4 = 114.05$	$N_4 = 12$

of Burt and Lewis (1) are presented<sup>1</sup> in Table 1. Evidently the four *remedial instruction methods*—Alphabetic, Kinesthetic, Phonic, and Visual—have significantly different effects at the 1% level.

However, typically the investigator may not wish to stop with the analysis of Table 1. He may wish to go on to examine Table 2. That is, for each significant over-all  $F$  emerg-

could be attributed in part to the difference between the Alphabetic and Phonic,  $T_1 - T_3$ ; or possibly to the contrast between the Kinesthetic and Visual,  $T_2 - T_4$ . In fact, going beyond a simple difference of two

<sup>2</sup> If the interaction between *remedial instruction methods* and *original teaching methods* were significant, then an investigation of the  $4 \times 4 = 16$  treatment combination would be in order; i.e., the investigator would want to examine the significance of the *remedial effects* for each level of *original teaching effects*, and vice versa.

<sup>1</sup> The data are also summarized as an example in Walker and Lev (8), Chap. 14.

treatments, the investigator might even wish to compare the Alphabetic and Phonic methods, *jointly*, with the Kinesthetic and Visual, *jointly*, i.e.,  $T_1 + T_3 - T_2 - T_4$ .

Two possible cases arise in the selection of such experimental comparisons for statistical testing: first, the *a priori* situation, i.e., where the comparisons have been planned in advance of the experiment; second, the *a posteriori* (or "post-mortem") situation, i.e., where the comparisons have not been pre-planned, but are suggested by an examination of the data. Only for the first case are the tabular  $F$  and  $t$  values appropriate, since these values are based on statistical theory which assumes *randomly* selected observations. Therefore, the conventional  $t$  or  $F$  test may be misleading in the second case (which is the more typical and acutely frustrating case to the investigator), where the comparisons are not determined solely by the nature of the experiment and design but are dependent upon the data.

Special methods are needed in this second case because the confidence levels are upset when the results of the investigation are used to decide on the comparison to be made. The most notorious example of making tests of significance on hypotheses suggested by the data is that of waiting until the experiment is completed and then selecting the highest and lowest means for analysis by an ordinary  $t$  test (or the equivalent  $F$  test for 1 degree of freedom in the numerator). The blunder in this procedure lies in the fact that the sampling distribution of such mean differences (between largest and smallest means) is more variable than the conventional distribution. And, as indicated by Cochran and Cox (2), in general the error variances for such "post-mortem" comparisons are

often considerably larger than those error variances appropriate to comparisons which have been hypothesized for test independently of the current data. The direction of the bias is clear. By using conventional (case one) test procedures in this unconventional (case two) situation, too small an error variance is used and the investigator will tend to obtain a large  $t$  or  $F$  value, yielding a spuriously significant comparison.

Various routes are open to an investigator desirous of fully examining his data yet aware of the danger of an erroneous inference. The following procedure has both the virtues of simplicity and conservatism to recommend it. Test the comparison by the ordinary numerator-one-degree-of-freedom  $F$  test (or equivalent  $t$  test). (a) If the comparison is not significant at the stipulated confidence level by this test, then the investigator can be sure of nonsignificance. (b) If the comparison is apparently significant by this test, the significance is possibly spurious. The technique illustrated below for dealing further with this (b) situation is due to Scheffé (7). Essentially, one simply uses the same computed  $F$  ratio but alters the apparent size of the critical region, conservatively.

For example, consider the "post-mortem" testing of the null hypothesis that the Alphabetic and Phonic effects are identical. For this one degree of freedom comparison,  $C: T_1 - T_3$ , the sum of squares<sup>3</sup> is

$$\begin{aligned} SS_C &= \frac{(T_1 - T_3)^2}{N_1 + N_3} = \frac{(1267.3 - 1247.2)^2}{12 + 12} \\ &= 16.83. \end{aligned}$$

The mean square for this comparison

<sup>3</sup> Walker and Lev (8, p. 356), give computing instructions for finding the SS due to any comparison.

is therefore  $MS_C = SS_C/1 = 16.83$ . Finally, the observed  $F$  ratio is

$$F = MS_C/MS_{Error} = 16.83/31.5 = .53$$

where the denominator mean square, with  $n_2 = 32df$ , is taken from the analysis of variance in Table 1. The tabular 1%  $F$  value for  $n_1 = 1$  and  $n_2 = 32df$  is  $F_{1\%,1,32} = 7.50$ . Hence this comparison does not reach the 1% significance level of the conventional test. A fortiori, it is not significant, i.e., it would surely fail to reach significance if the appropriate 1% level were available. This illustrates situation *a* above.

It should be noted that instead of performing the test of significance by means of the  $F$  ratio, a general alternative is to use the corresponding  $t$  statistic. They are entirely equivalent since  $F_{1,n_2} = t_{n_2}^2$ , i.e.,  $F$  with  $n_1 = 1$   $df$  for the numerator and  $n_2$   $df$  for the denominator is equal to the square of  $t$  with  $n_2$   $df$ . As illustration, the present comparison  $C: T_1 - T_3$  can be equivalently expressed as a difference of means,  $C: M_1 - M_3$ , instead of as a difference of totals. Then  $t = (M_1 - M_3)/SE_{M_1 - M_3}$ , where the standard error<sup>4</sup> is given by

$$SE_{M_1 - M_3} = \sqrt{MS_{Error}(1/N_1 + 1/N_3)} \\ = \sqrt{31.5(1/12 + 1/12)} = 2.29.$$

Hence

$$t = \frac{105.61 - 103.93}{2.29} = .73,$$

the square of which (.5329) is approximately  $F = .53$ . Whether  $t$  or  $F$  is used in such applications is a matter of taste, except in the case of a one-sided hypothesis where, of course, the  $t$  statistic is employed since the sign of the comparison is funda-

mental. Jones (4) discusses the rationale of unilateral comparisons.

As a second example of a "post-mortem" statistical test, consider the hypothesis that the effects of the Alphabetic and Phonic methods, *jointly*, is identical with the *joint* effect of the Kinesthetic and Visual treatments. For this one degree of freedom comparison,  $C: T_1 + T_3 - T_2 - T_4$ , the sum of squares is

$$SS_C = \frac{(T_1 + T_3 - T_2 - T_4)^2}{N_1 + N_2 + N_3 + N_4} \\ = \frac{(1267.3 + 1247.2 - 1347.2 - 1368.6)^2}{12 + 12 + 12 + 12} \\ = 844.20.$$

The mean square for this contrast is therefore  $MS_C = 844.20$  and the observed  $F$  ratio is  $F = 844.20/31.5 = 26.8$ . Since 26.8 exceeds the 1% point of 7.50, then this comparison is apparently significant. However, this example illustrates situation *b* above; i.e., the apparent significance of this contrast may be spurious. The Scheffé modification of the significance region is therefore in order. In general, at the  $\alpha\%$  significance level, it consists of using as significance point,<sup>5</sup> the value  $(k-1)F_{\alpha,k-1,n_2}$  instead of  $F_{\alpha,1,n_2}$ . In this expression,  $k$  denotes the total number of treatments in the group of treatments from which the comparison has been constructed. Thus here the value  $(4-1)F_{1\%,4-1,32}$ , or  $3F_{1\%,3,32}$  which is (3) (4.46) = 13.38, would be used instead of  $F_{1\%,1,32} = 7.50$ . Since the observed  $F$  of 26.8 exceeds even this modified 1% significance point of 13.38, the reality of the difference be-

<sup>4</sup> Lindquist (5, p. 14), gives computing instructions for evaluating the standard error for any comparison.

<sup>5</sup> When the comparison of interest is a simple difference between two means, a more sensitive test due to Tukey may be employed. Tukey's test requires tables of the Studentized range instead of the  $F$  table only. An example is to be found in Scheffé (6).

tween the effects of Alphabetic-Phonic and Kinesthetic-Visual seems assured. Evidently, the significance region is altered conservatively by the Scheffé method, i.e., a larger observed  $F$  is needed to obtain signifi-

cance at a stated level when the comparison is devised a posteriori (and so capitalizes on possibly fortuitously large effects) instead of a priori. This is in accord with the considerations mentioned earlier.

#### REFERENCES

1. BURT, C., & LEWIS, R. B. Teaching backward readers. *Brit. J. educ. Psychol.*, 1946, 16, 116-132.
2. COCHRAN, W. G., & COX, G. *Experimental designs*. New York: Wiley, 1950.
3. JOHNSON, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.
4. JONES, L. V. Tests of hypotheses: one-sided vs. two-sided alternatives. *Psychol. Bull.*, 1952, 49, 43-46.
5. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.
6. SCHEFFÉ, H. An analysis of variance for paired comparisons. *J. Amer. Statist. Ass.*, 1952, 47, 381-400.
7. SCHEFFÉ, H. A method for judging all contrasts in the analysis of variance. *Biometrika*, 1953, 40, 87-104.
8. WALKER, H. M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

Received July 6, 1954.



## THE INFANTILE DISORDERS OF HOSPITALISM AND ANACLITIC DEPRESSION<sup>1</sup>

SAMUEL R. PINNEAU<sup>2</sup>

University of California

Within the last ten years, psychogenic disturbances in infancy have received increasing recognition (34). In this connection Spitz (43, 44, 45), Ribble (38, 39), and Fischer (16) have especially emphasized the importance of the mother-child relationship, and their reports have been increasingly accepted and quoted by many. At least two major textbooks in general psychology have given considerable space to these authors' works (25, 41). Their works receive even greater attention and acceptance in Maslow and Mittelmann's recent revision of their text for abnormal psychology (34), where they constitute approximately half of the chapter on disorders of infancy. A favorable reception is also apparent in articles by Geleerd (18), by Hartmann, Kris, and Loewenstein (21), and by Bowlby (9). Kris' evaluation of these writers' works appears to be representative of this school: "... it seems appropriate to state how much we owe to Margaret Ribble's own investigation ... and to the long set of investigations on the consequences of the early institutionalization of the child, to which R. Spitz has contributed so decisively ..." (30, p. 31). Less favorable, or even skeptical, appraisals may be found in reports by Jones (29), by Dennis (14), and by

Orlansky (35). The increasing recognition given these studies and the somewhat contradictory evaluations of them indicate the need for a more complete and critical consideration of them.

As the writer has evaluated the writings of Ribble elsewhere (36), they will not be considered in the present review. Primary consideration will be given to the investigations which Spitz reports (43, 44, 45, 46, 47) as these are the most extensive and as they have been reported in the greatest detail. A brief summary of his investigations will be presented before we attempt to evaluate them. An evaluation of Fischer's work may be found in a separate section at the end of this article.

### THE SPITZ PAPERS

The Spitz papers which will be considered in this article are those which have been published in the first six volumes of *The Psychoanalytic Study of the Child*. The first of these, entitled "Hospitalism. An Inquiry Into the Genesis of Psychiatric Conditions in Early Childhood," was published in 1945 in Volume I of this series (43). In this paper he contrasts the development of infants in two institutions which he names "Nursery" and "Foundling Home." These will be described in some detail in the next section. The report deals specifically with an analysis of the data in terms of the children's development as measured by the Hetzer-Wolf tests, a comparison of the background of the two groups, a com-

<sup>1</sup> The writer wishes to acknowledge the criticisms and suggestions of Harold E. Jones, which have been invaluable in the writing of this article.

<sup>2</sup> Part of the work on this review was done while the writer was on the staff of the University of Oregon.

parison of their physical development, and a comparison of their care in the two institutions in terms of housing conditions, food, clothing, medical care, personnel, restrictions on activity, and amount of contact with their true or substitute mothers.

Spitz's second and third articles were published in 1946 in the second volume of this series. The second article, "Hospitalism. A Follow-up Report on Investigation Described in Volume I, 1945," deals chiefly with a description of the development of the Foundling Home infants in the two years subsequent to the original study (44). These descriptions were based on data collected by an investigator who visited Foundling Home at four-month intervals during this time. Only 21 of the original group of 91 infants were still available for study at the end of this time. He contrasts the development of these children with the development of 122 infants from the other institution, Nursery. As only 69 infants were present in Nursery in the original study (43), 53 new cases had been added to the sample for this study.

His third article is entitled, "Anaclitic Depression. An Inquiry into the Genesis of Psychiatric Conditions in Early Childhood, II" (45). The term *anaclitic depression* refers to a psychiatric syndrome of depression which Spitz observed in some of the infants studied in the institution called Nursery. This syndrome was evident in severe form in 19 of the infants and in mild form in 26 of them. All cases manifesting the disorder had been separated from their mothers, but not all who underwent separation developed the syndrome. (The number of children who were separated from their mothers was not given.) He contrasts the reaction of the 45 Nursery children manifesting anaclitic depression with the develop-

ment of the children in Foundling Home who were separated from their mothers prevalently in the sixth month. In addition he includes a review of the literature on the concept of early depression.

His fourth article, "Autoerotism. Some Empirical Findings and Hypotheses on Three of its Manifestations in the First Year of Life," was published in 1949 in Volume III-IV of *The Psychoanalytic Study of the Child* (46). The three manifestations of autoerotism investigated were "rocking," genital play, and fecal games. He relates these to each other and to the intensity of the mother-child relationship. He compared the incidence of these activities in three groups of subjects, those reared in Nursery, Foundling Home, and family homes. The main body of the investigation was based on 170 of the 190 infants now included in the Nursery sample.

The fifth and final article with which we will be concerned was called "The Psychogenic Diseases in Infancy: An Attempt at their Etiologic Classification" (47). It was published in 1951 in the sixth volume of this series. This paper presents Spitz's attempt to set up a classification of psychogenic diseases of infancy. This report adds little to our knowledge of the growth and development of the children in Nursery and Foundling Home.

#### NURSERY AND FOUNDLING HOME

Before turning to some of Spitz's broader conclusions, it would seem desirable to consider in a general way the circumstances which led to his subjects being reared in these institutions, the background of these children's parents, the physical characteristics of the institutions and the care received by the children while there.

Nursery is a penal institution for delinquent girls. In that institution, the girls who gave birth to a child while there, cared for their infants until they were approximately a year old. While the Nursery infants are generally described as being cared for by their mothers throughout the first year, it appears in a later report by Spitz (45) that a considerable proportion of them were separated from their mothers for a period of three months beginning at 6-8 months of age, for reasons not indicated. When a child had to be separated from his mother, a pregnant girl or another mother cared for him. The mothers were supervised by a head nurse and three assistants. Despite the fact that he describes most of the mothers as socially maladjusted, feeble-minded, psychically defective, psychopathic, or criminal, Spitz reports that in Nursery the mother "gives the child everything a good mother does and, beyond that, everything else she has" (43, p. 65). The children were kept in glass-enclosed cubicles until they were six months of age, at which time they were transferred to rooms in which there were a number of children. Toys were almost always present and there was a feeling of warmth and friendliness about the institution because the mothers spent a great deal of time carrying their children, tending them, feeding them, playing with them, and chatting with each other with their babies in their arms.

In contrast to Nursery is Foundling Home. The children were taken to that institution because their mothers were unable to care for them outside of the institution. It appears that inability to support the children was the main reason that the mothers were unable to care for them and not because they were socially maladjusted or abnormal in any way. The infants were breast-fed by their

mothers who were present in the institution for several months but, for reasons not given by Spitz, the mothers seem to have had little to do with the care of their children. The children in Foundling Home were cared for primarily by five to eight nurses who were described as unusually motherly, baby-loving women. Each of these nurses cared for eight or more children and, apparently as a result, the children lacked human contact most of the day. These children also lived in cubicles, but in this case they were enclosed on only three sides. Bed sheets were hung over the sides of the crib so that only rarely were the infants able to see what was going on in the ward about them. Foundling Home was poorly provided for financially as compared with Nursery and, when Spitz first went to the institution, about the only objects that the children had to play with were their own hands and feet.

#### SPITZ'S CONCLUSIONS

From his study of these subjects, Spitz concludes that those infants who are separated from their mothers for over six weeks tend to develop psychogenic disorders. We will use the term *hospitalism* to cover the syndrome of symptoms which characterized the infants of Foundling Home, in keeping with his discussion of this syndrome in his initial article (43). As previously stated, the syndrome manifested by the children in Nursery who were separated from their mothers will be termed *anacletic depression*. Spitz makes it evident in an article (47) published some five years after his initial reports (43, 44, 45) that *hospitalism* is merely an exaggerated form of the disorder of *anacletic depression*; however, his original descriptions of the symptoms differ somewhat, and for the purposes

of this review they will be treated separately.

Two major symptoms characterized the infants of Foundling Home. One of these, and the most prominent, was a drop in the developmental quotient (DQ) as measured by the Hetzer-Wolf baby tests. The second most prominent was a change in the infants' reactions to strangers during the last third of the first year. In contrast to their usual reaction to strangers, the children's behavior now varied from blood-curdling screams to an extreme friendliness to strangers combined with an anxious avoidance of inanimate objects (43). Other symptoms of the hospitalism syndrome include retardation in skeletal development, in ability to sit and walk, and in ability to develop social skills (44). These symptoms became increasingly marked with the length of the period of separation from the mother. He concludes that these symptoms are the result of the child being separated from his mother. While he notes that the children were deprived in other ways also, he states, "The presence of a mother or her substitute is sufficient to compensate for all the other deprivations" (43, p. 68).

Spitz termed the syndrome "anaclitic depression," because he considered that the clinical picture was similar to that found in depression in adults. Among the symptoms shown by the infants were the following: There was a drop in the DQ as measured by the Hetzer-Wolf baby tests. They developed a weepy, apprehensive, sad behavior in contrast to their previously happy and outgoing behavior. They lay or sat with wide-open, expressionless eyes and frozen immobile faces. They looked as if they were in a daze and apparently were not perceiving what went on in their environment. This behavior

was accompanied in some by autoerotic activities in the oral, anal, and genital areas. Not every symptom was apparent in every case and it seemed that, at least in some, first one symptom dominated and then another. He comes to the same conclusion in this study that he came to in his study of Foundling Home: The cause of the symptoms is the separation of the child from his mother.

We will turn now to an evaluation of his studies to determine if his conclusion is warranted by the data he presents.

#### DIFFICULTIES INVOLVED IN AN EVALUATION OF SPITZ'S STUDIES

A number of features of Spitz's reports on his study of the Nursery and Foundling Home infants make an evaluation of his findings difficult. Spitz gives practically no information as to the time at which the studies took place or as to the location of the institutions; however, some information can be gleaned from his various reports (43, 44, 45, 46, 47). In a footnote to the first report, published in 1945, he notes that his approach to the problem of hospitalism was "... mapped out and begun in 1936 ..." (43, p. 56). In a report (44) written June 12, 1946, he noted that the institution termed Foundling Home was first visited two years previously, and that the study of the other institution, Nursery, "... now covers a period of three-and-a-half years ..." (44, p. 116). Judging from this information, it appears that the observation of the children began in late 1942 or early in 1943. While stating that the institutions are in two different countries in the Western Hemisphere, he does not designate the countries more specifically and, according to the magazine *Time* (50), has refused to do so. From his references it was possible to determine

that Nursery is an institution for delinquent girls located somewhere in New York state (46, 49); however, one cannot from his writings locate Foundling Home more specifically than south of the Rio Grande. The secrecy preserved concerning social and geographic areas served by these institutions is one of the factors which make it extremely difficult to evaluate the Spitz results.

Some idea regarding the magnitude of the investigation can be gained from his third report. He says regarding children in Nursery,

These 123 infants stayed in the nursery from their fourteenth day to the end of their first year and in a few cases up to their eighteenth month. No selection was made in the infants observed. We invariably tested and followed each child admitted to the nursery up to the day when it left. The observations took place at weekly intervals, and totalled approximately 400 hours for each child (45, p. 317).

This statement would seem to imply individual observation on each child. Spitz does not reveal the size of the population in Nursery at a specified time, except to say that when the study began there were 69 infants. On the basis of an average stay of 14 months, this would seem to imply a requirement of 460 hours of observation per week, and a staff of 12 persons. In a study of this magnitude, we would usually expect to have the staff named, and some information concerning their training and qualifications, but this is another point on which Spitz remains silent.

#### *The Samples*

One of the most important considerations in comparing two groups of subjects is an analysis of the respects in which they are different and comparable. As Spitz's conclusions are based in part on the differences between the two groups at certain ages, it would seem appropriate to consider

what differences would be expected on the basis of what we know about the backgrounds of the subjects. It is important that individual cases were not excluded in the study, that it was planned so as "... to embrace the total population of both [institutions] (130 infants)" (43, p. 56). That selection was absent in all the reported investigations is apparent for he says in his fourth report, "As in all our investigations, the unselected total sample of the children present in the institution was observed by us and used for our study" (46, p. 88). It is regrettable that he did not note the absence or frequency of congenital abnormalities in the two institutions, the birth condition of the infants, incidence of syphilis in the mothers, or other health conditions which might be relevant, and that he did not note what if any effect such factors might have had on his findings. However, these groups can better be compared after they have been discussed separately.

#### *The Nursery Sample*

As noted earlier, the institution referred to as Nursery is a penal institution for delinquent girls (43). Some of the women were pregnant on admission. After delivery and the lying-in period, the mothers assumed the care of the infants who remained in the Nursery until the end of their first year. Spitz describes the mothers as follows:

The mothers in Nursery came there because of a failure in social adaptation. In a large percentage of the cases this maladaptation is not severe, consisting mainly in sexual indiscretion at the wrong age. ... We suspect that the difference between the mothers in this institution and other mothers of an urban background is one based on cultural attitudes of their immediate environment and on the diversities of their economic status (46, p. 98).

Apparently Spitz continued to add cases to his sample. In his first article



(43) Spitz reports that the total number in the Nursery sample was 69, in the second article (44) 122, in a third (45) 123, and in a fourth (46) 196. There is no indication as to the ages at which the new cases were added. It is not until the third article that the group is described as both white and colored (77 and 46 cases, respectively). Since such an obvious sampling detail was omitted, one may wonder if there were other sampling details, relevant to interpreting his findings, which were also omitted.

### *The Foundling Home Sample*

It has been indicated that children are placed in Foundling Home because their mothers are unable to provide for them. Regarding their background Spitz says:

A certain number of the children housed have a background not much better than that of the Nursery children; but a sufficiently relevant number come from socially well-adjusted, normal mothers whose only handicap is inability to support themselves and their children (which is no sign of maladjustment in women of Latin background) (43, p. 60).

While this limited amount of information makes an adequate evaluation of the children's background impossible, the mothers' inability to support their children does raise several questions. Thus, is their inability to support their children the result of being of the lowest socioeconomic group, of having limited mental abilities, or of both? Why did the responsibility of supporting the children fall to the mothers rather than to the fathers? What were the handicaps or incapacities of the fathers that made this necessary? Is it reasonable to expect that one would find in *any* foundling home that "... the children are of an unselected urban ... background" (43, p. 58), as Spitz maintains for this group? Certainly his statement that

the mothers were unable to support the infants would seem to indicate that the sample was biased in its socioeconomic level.

As in the case of the Nursery group, the actual size of the Foundling Home sample is not easy to determine. Spitz (43) states that the study embraced the total population of children in Foundling Home, and in the subsequent table (43, p. 57) gives the number of children present as 61. Later in the same article he states that there were 88 children in Foundling Home up to 2½ years of age (43, p. 59). In a subsequent article he refers to 91 children under 3 years of age (44, p. 114).

Bowlby states that Spitz, in his first article, gives "explicit data" showing that the mother-separated and control groups "... are of a similar social class and, as nearly as possible, spring from similar stock" (9, p. 20). However Spitz, in comparing the background and heredity of the two groups, contends that his data show a *marked advantage* for the Foundling Home children (43, p. 61, 66). The author can find no support for either of these contentions in Spitz's reports. In fact from the limited information available, it appears that the Nursery group may have the advantage.

According to Spitz the difference between the Nursery mothers and other mothers of an urban background is to be accounted for in terms of differences in their cultural backgrounds and in the diversities of their economic status (46, p. 98). If true, there would appear to be no basis on which to maintain that the Foundling Home children have a superior heredity since previous considerations indicate that they come from the lowest socioeconomic group.

The behavioral differences between the two groups could represent, in

various combinations, differences in racial extraction (45) and racial mixture (43, 45), negative hereditary selection (43), and congenital defects, as well as differences in cultural factors and economic status (46). It would appear that only very tentative conclusions as to the importance of the mother-child relationship should be based on differences in behavior of two groups in which there are so many other possible factors which may affect developmental data.

### *The Nature of the Study*

It has been noted that the conclusions which Spitz draws in his reports are based, in part, on a comparison of the Nursery with the Foundling Home infants. Spitz also draws conclusions from comparisons of the age-trends of a given group with other groups and with developmental norms.

It seems to the writer that anyone who reads Spitz's articles is likely to infer from his statements that his reports are based on longitudinal data. However, there is question as to the extent to which this is true.

First let us determine if Spitz implies that a longitudinal method was used. His description of the Foundling Home children's development follows: "Their Developmental Quotient on admission is below that of our best category [a group of children from 'professional homes'] but much higher than that of the other two [the Nursery infants and a group of infants from an isolated fishing village]. The picture changes completely by the end of the first year, when their Developmental Quotient *sinks*<sup>3</sup> to the astonishingly low level of 72" (43, p. 58) and, "By the end of the second year the Developmental Quotient *sinks*<sup>4</sup> to 45 . . ." (43, p.

70). Elsewhere he states, "The children . . . [in Foundling Home] though starting at almost as high a level as the best of the others, had spectacularly *deteriorated*"<sup>5</sup> (43, p. 59). He maintains that his data show the "comparison of the development in Nursery and Foundling Home" (43, p. 67, 69). This emphasis on *change* in score and on *development* would seem to imply that on the average the DQ's of the *same* group of children deteriorated from a score of over 130 during their second month to a DQ of 72 by the end of their first year, and to 45 by the end of the second year.

If the reader has drawn this conclusion, it will become apparent that he has made an error.<sup>6</sup> Spitz says (43, p. 59) that at the beginning of his study there were 88 children in Foundling Home below 2½ years of age, and only 45 of these were under 1½ years of age, i.e., there were on the average only 2½ children at each month of age from birth to 18 months and less than 4 at each month of age from 12 to 24 months. Hence, if each case was followed longitudinally, that portion of his graphs covering the first year of development could not have been based on more than half of his subjects, and that portion covering the first six months, the period of most rapid decline in developmental quotient, could not have been based on more than a fifth of his cases. It would be impossible for the graphs to represent the change of performance for a constant sample of 88 children. Rather it appears that his graphs present the average DQ of different children at each age. That this is the correct interpretation becomes evi-

<sup>3</sup> Italics mine.

<sup>4</sup> Italics mine.

<sup>5</sup> Maslow and Mittelmann appear to make this error in their summary of Spitz's work (34).

<sup>6</sup> Italics mine.

dent when we consider his follow-up study written in July, 1946 (44). He states that he *first* visited this institution two years prior to this report and he also states that the follow-up study had been going on for two years. From these statements one must conclude that the original study must have consumed very little time—so little time that he could still state after a two year follow-up study that the youngest of the children still in the study was two years of age. While he does not give this child's age as two years zero months, this is implicit since in the same sentence he gives the oldest child's age in years and months (four years and one month) (44, p. 114). Inasmuch as the time required for the original study was so brief, and inasmuch as *tests were not administered to the children in the follow-up study* (44, p. 113), the graphs he presents for Foundling Home infants could not be based on longitudinal data. Thus it becomes obvious that conclusions regarding the development of the Foundling Home infants, as measured by the Hetzer-Wolf tests, must be restricted by the cross-sectional nature of the study.

Let us turn from the Foundling Home infants to the Nursery sample. As the Nursery was not a new institution which took in its infants at one time and of the same age (44), the 69 infants included in the first report undoubtedly varied in age between birth and twelve months. If one assumes that they are evenly distributed with respect to age, there would be 5.75 children at each month. Presumably about three-quarters of the 69 children were over the age of three months at the time the study began and hence were not observed and tested during their first three months. In addition it would mean that about half of the

total sample was not observed and tested during their first six months. Hence, the graphs based on this sample of children must to a large extent represent different children at each age and not the change in performance of 69 children from one age to the next, as would seem to be implied by various statements and by the labeling of the graphs.

At a later time additional cases were added to the Nursery sample. In the article entitled "Anaclitic Depression" he stated, "In the course of a long term study of infant behavior in a nursery where we observed 123 unselected infants, *each*<sup>7</sup> for a period of twelve to eighteen months . . ." (45, p. 313). Of course all of these 123 children could not have been observed " . . . during the whole of the first year of their life . . ." (45, p. 314), inasmuch as approximately 50 of the children in the original sample of 69 cases were over 3 months of age when the study began. There can be little doubt that this group of 123 infants included these 50 children, inasmuch as he says in his fourth article, "As in all our investigations, the unselected total sample of the children present in the institution was observed by us and used for our study" (46, p. 88). If 123 children represented all the children present in the institution at the time that this report was made, one would think that a sizable number of these would have been under twelve months of age. If so, these infants, together with those of the initial sample that it was impossible to observe during the whole of their first year, probably constitute more than half of the 123 cases. Thus, probably more than half of the total sample could not have been observed " . . . during the whole of the first year of their life . . ." (45, p. 314).

<sup>7</sup> Italics mine.

These considerations make it difficult to evaluate such statements by Spitz as that in which he says that all 123 children had been observed "... each for a period of twelve to eighteen months ..." (45, p. 313).

#### THE BEHAVIOR AND DEVELOPMENT OF SPITZ'S SUBJECTS

According to Spitz, the most prominent feature of the disorder manifested by the Foundling Home infants is severe developmental retardation as determined by the Hetzer-Wolf baby tests (43). The second most striking feature is a change in the child's reaction to strangers when he is between eight months and a year of age. For the disorder of anaclitic depression among the Nursery group, he lists three groups of symptoms—the static, genetic, and quantitative signs. So far as the static and genetic signs are concerned, Spitz does not indicate how they were assessed. In addition, no comparisons are made with the frequencies with which these signs appear in normal control samples. The quantitative signs are based on performance on the Hetzer-Wolf baby tests. He states that the subjects "... show a gradual drop of the developmental quotient; this drop progresses with the progression of the disorder" (45, p. 328). Of the three, this symptom alone appears to be well-defined, and it alone depends on a standardized procedure of known reliability.

Before turning to this special aspect of the infants' development and to the measuring instrument employed, attention should perhaps be directed to other characteristics of their development. Spitz states that the Foundling Home children do not turn from back to side, even as late as the seventh month. In this particular respect these infants are

very retarded in comparison with developmental norms.<sup>8</sup> The fact that they "... lie supine in their cots for many months ..." (43, p. 63) cannot be attributed to separation of the infants from their mothers, as separation took place most prevalently during the sixth month (45, p. 331). Spitz gives another interpretation to this finding, namely: "... a hollow is worn into their mattresses... this hollow confines their activity to such a degree that they are effectively prevented from turning in any direction" (43, p. 63). This description, if the correct one, is hardly in accord with his other descriptions of the infants' care and of the physical conditions of Foundling Home: "... they were adequately cared for in every bodily respect ..." (47, p. 271). "... hygiene and precautions against contagion were impeccable ..." (43, p. 59). "As regards food, housing, clothing, hygiene, the conditions were comparable to those encountered in 'Nursery.'" (46, p. 94). In addition, most of the mothers were apparently present to help care for the infants until they were approximately six months of age. Furthermore, Spitz, states that "Foundling Home is visited by the head physician and the medical staff at least once a day, often twice, and

<sup>8</sup> Shirley, in her study of 25 babies, notes that "A few babies turned completely from the back to the side; such reactions were observed in six babies between the ages of 1 and 11 days ..." (42, p. 44) and "Five babies were observed to roll in this way [from the prayer position\* to the back] before the 12th week" (42, p. 44); Gesell gives 3 months as the age placement for "Rotates body from dorsal to side position ..." (19, p. 129); Dennis (13, 15) gives 27 weeks as the median age for rolling from supine to prone, but gives as the range, 7 to 34 weeks; and Halpern states that "The healthy full-term infant is usually able to roll from the front of his body to the back and vice versa beginning in the fifth month" (20, p. 219).

during these rounds the chart of each child is inspected as well as the child itself" (43, p. 62). It would appear that we may be somewhat skeptical about the general physical environment and care of these infants if they are *actually* confined in holes in their beds so deep that they can not turn in any direction.

Spitz says that "As soon as the babies in Foundling Home are weaned, the modest human contacts which they have had during nursing at the breast stop, and their development falls below normal" (43, p. 66). While "The time when the children in Foundling Home are weaned is the beginning of the fourth month" (43, p. 66), we note that "... the separation from the mother took place beginning after the third month, but *prevalently*<sup>9</sup> in the *sixth*<sup>10</sup> month" (45, p. 331).

Let us consider now the time-location of the "decline" in DQ according to his graphs (43, p. 69; 47, p. 272). During the second month the average DQ was approximately 131 for Foundling Home; in the fifth month approximately 88; in the sixth month approximately 76; in the seventh month approximately 75, and at a year approximately 72. Thus we see that there was a difference of approximately 59 DQ points between the children in their second month and the children at a year of age. However, 43 points of this drop occurred by the time the children were tested in their fifth month and hence while the majority of the mothers were still present. The difference between the children in their fifth month and those in their sixth month, the month in which separation from the mother was most prevalent (45, p. 331), is 12 points. The difference between the children in their sixth

month and the children at a year is only approximately four points.

The data just reviewed support Spitz's contention that marked retardation characterizes the Foundling Home infants; however, the information given by Spitz, rather than supporting his hypothesis that the retardation is due to separation of mother and child, indicates that it was in evidence before the separation.

As Spitz places a great deal of weight on the data derived from the Hetzer-Wolf baby tests, a careful evaluation of them is necessary.

#### *The Tests Used in the Study*

As indicated earlier (43, p. 72), the most prominent feature of the disorder of "hospitalism" is severe developmental retardation as measured by the Hetzer-Wolf baby tests. It plays an equally important role in the diagnosis of anaclitic depression (45). As his reference to these tests he gives Hetzer and Wolf's article "Baby Tests," published in 1928 (23). Since he always makes reference to the originally published tests rather than to the revision by Frankl and Wolf (11), there can be little doubt that the original scale was used.<sup>11</sup> (These tests are frequently referred to as the "Bühler scales" [22, 27, 32].)

Inasmuch as the baby tests were originally standardized on Viennese children, the norms may or may not be applicable to children in other cultures. Perhaps some consideration

<sup>11</sup> The above assumption may or may not be justified inasmuch as he says regarding the tests, "They provide . . . quantifiable data on six distinct sectors of personality . . ." (43, p. 55), and only four "sectors of personality" were indicated in the original form of the tests. Six were designated in the revised tests, but the revision was made by Frankl and Wolf and it would appear inaccurate to designate these latter tests as the "Hetzer-Wolf baby tests."

<sup>9</sup> Italics mine.

<sup>10</sup> Italics mine.



should be given to the original standardization before studies from other cultures are considered.

#### *Standardization*

According to Hetzer and Wolf (24, p. 203), the tests were standardized on 20 children at each month level. The subjects were children in the Reception House for Children in Vienna (24, p. 203). Most of the children of the Reception House came out of family care to spend a three-week quarantine period before being placed either with a foster mother or in another institution (10, p. 6). Hence these children were themselves subject to maternal deprivation prior to testing. According to Bühler, "Children used in these tests were largely taken from the poorer population of Vienna" (10, p. 89), and hence as Herring notes (22), the norms are not representative of Viennese infants in general.

#### *Studies on Infants of Other Cultures as Related to Spitz's Findings*

*The average DQ.* In a search of the literature, three studies, all cross-sectional, were found which are relevant to the use of the test with children from other cultures. All three cast doubt on the accuracy of the norms, at least for children of other cultures. The first of these studies was carried out by McGraw in her comparative study of a group of southern white and Negro infants (32). "About fifty per cent of the babies [in Tallahassee, Florida] born within the year—both white and colored—figured as subjects in this comparative study" (32, p. 28). "Although the subjects were selected at random and the results for both white and colored yield [for the total sample] a normal distribution in terms of 'Developmental Quotient,' the educational levels of the parents

tend toward the upper grades, high school, and college levels for both white and colored . . ." (32, p. 33-34). McGraw reports the average DQ for each age group on the Hetzer-Wolf test.<sup>12</sup> These are presented in Figure 1. As is apparent in the figure, the average DQ tends to decrease with age for both of McGraw's groups.

The second of these studies was that carried out by Ruth Hubbard on the reliability and validity of the scale (27). She does not give average DQ's by month for her 78 babies ranging from 1 to 20 months, but she does conclude from her study that the scale is well worth the efforts of a thorough standardization on American infants. She notes a number of disadvantages of the scale including " . . . the wide scatter of successes necessitating a testing period of an hour in many cases, the hiatus of a month between the eleven- and twelve-month levels for which no tests are provided, the number of frustrations in the series, and the inadequate standardization of administration and scoring" (27, p. 382).

The third study was carried out by Hsu with Peiping babies (26). While there were only twenty-six infants in his study, the trend of his averages are much like those of McGraw. Hsu concludes, "It is apparent that the Viennese scale rated the Peiping babies too high at the lower age level and too low at the upper age level" (26, p. 217).

<sup>12</sup> It should be noted that certain changes were made by McGraw in the test material and in scoring; however, she notes that in the English translation Rowena Ripin adopted a scheme of scoring not very different from that employed by Bühler and her associates. In her monograph McGraw notes a number of features of the test which she considers to be " . . . outstanding frailties in the scale as a means of measuring the development of infants" (32, p. 65-67).

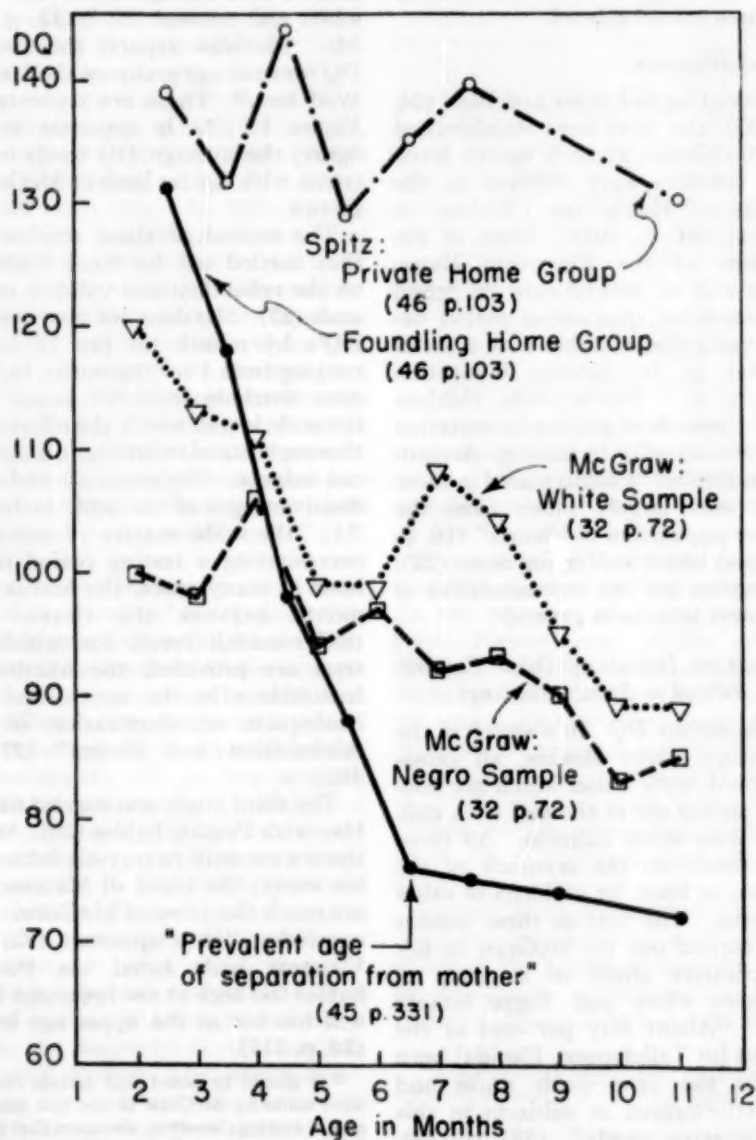


FIG. 1. MEANS OF DEVELOPMENTAL QUOTIENTS WITHIN THE FIRST YEAR FOR THE HETZER-WOLF "BABY TESTS"

Spitz's curve for the Foundling Home infants (43, p. 69), is included in Figure 1. The average DQ's of his Foundling Home infants, like the average DQ's in the studies of McGraw and Hsu, decrease with age during the first year. An analysis of this figure supports the contention of Hsu that the Viennese infant scale rates children too high in the early months and too low in the later months. It is regrettable that Spitz was apparently unaware of this problem.<sup>18</sup> However, Spitz's Nursery group did not show a decline in DQ with age. Discussion of the results from this group will be deferred until a later section.

The possibility that the Frankl-Wolf revision of the Hetzer-Wolf baby tests was used in the study should not be overlooked. Two major studies of this scale were found in the literature. The first of these was by Ackerman (1) utilizing 25 children at each of four levels (7 months, 8 months, 9 and 10 months, and 11 and 12 months). The average DQ was 106.67. There was only a slight drop in the average DQ for these ages, the drop in means being 2.22 points from the tests at 7 months to those at 11 and 12 months. However, it should also be noted that during this period the differences in DQ of the Foundling Home children in Spitz's study were essentially of the same magnitude, being only three or four points lower at 12 months than they had been at 6 to 7 months. The study sheds no light on those months of most concern to us, those in which

Spitz's age differences were "significant." In contrast, we may consider Herring's study (22) on the reliability of the test. In her report on 114 subjects the mean DQ is given for each group. The same trend as was noted for the original scale (i.e., for the averages to be successively lower) is present, although it is not as marked. (The mean DQ's for 1, 2, 3, 5, 6, and 7 months were respectively 132.7, 128.1, 90.9, 115.5, 119.8, and 107.6.) This trend suggests that the scale, even when revised, is inadequately standardized and gives a spuriously high score at the earlier ages.

Spitz also used the Hetzer-Wolf baby tests on a group of 17 children raised in private homes of white-collar workers. From a close consideration of these homes, he felt that the relationship between the child and his mother was exceptionally good (46, p. 94). The average DQ's for these children for their first year were taken from his chart (46, p. 103) and are included in Fig. 1. One cannot determine from his chart or from the text of his article whether or not this curve is based on longitudinal data, nor can one tell how many of these children are represented at any one age.

It should be noted that Herring found no significant difference between the scores of infants from the three upper and three lower socioeconomic groups on the Frankl-Wolf tests (22). In essence, being from "private homes of white collar workers" does not predispose children of this age to have high scores. It is not apparent why Spitz's findings on these children reared in private homes are so out of line with Herring's report (22), with the findings of Bayley (7) that scores during early infancy have little relation to scores in the latter part of infancy, and with the

<sup>18</sup> Bühler maintains that in all studies called to her attention the median DQ has not differed significantly from a DQ of 100 (11, p. 88). If one considers the samples as a whole, without regard to age-level, the preceding studies would probably also agree with this contention; however, this would not raise the validity of the instrument as a measure of normal development at any given age level.

findings of Bayley and Jones (8) that scores within the first year are essentially unrelated to parents' socioeconomic, educational, and occupational level.

From the studies summarized in this section we may conclude that with most groups which have been tested, the Hetzer-Wolf tests give average DQ's which are considerably above 100 in the earlier months and considerably below in the later months. The Frankl-Wolf tests also appear to give average DQ's which are considerably above 100 in the early months. These considerations suggest that a considerable portion of the difference between the scores of the younger and older infants of Foundling Home could have been predicted on the basis of normative studies, and that these differences may be chiefly an artifact of the tests' standardization. Spitz's data obtained from his other groups are out of line with the majority of studies of the Hetzer-Wolf tests.

#### *Predictive Ability of the DQ*

Although the extent of the relationship between tests at different ages has not been determined for the Hetzer-Wolf baby tests, some information is available for the Frankl-Wolf revision. Herring (22) correlated the average of her subjects' scores on two successive tests at one month with their average on tests at five and six months and at nine and ten months. The correlations were respectively .288 for 22 subjects and .345 for 23 subjects. Neither of these is significant at the 5 per cent level. Her correlation of the average of scores at five and six months with the average of scores at nine and ten months for 22 cases was .448. While this correlation is significant at the 5 per cent level, it is still too small to be of much value for predictive purposes.

The preceding study indicates that one would do little better than chance in predicting from a single administration of the Frankl-Wolf test whether an infant's score four or five months later will be low or high. Since this test is a revision of the test used by Spitz, it would seem unlikely that the tests he used would have any greater predictive value. Support for this hypothesis, at least for longer time intervals, is found in Anderson's use of most of these items in his study of the predictive value of infancy tests (3). Other studies indicate that lack of predictiveness is a characteristic of infant scales; e.g., Bayley (7) in a complete analysis of the predictive value of an infant test, the California First Year Mental Scale (4), found that scores at six months of age are negatively related to performance at school age and beyond, and that the median correlation between performance at a year and tests subsequent to school age is only .23.

Spitz's reports would lead one to believe that a drop in score within the first year has some intrinsic value and that a low score by the age of one is of special significance. That a drop in score within the first year has little predictive value may be illustrated by considering those subjects in the Berkeley Growth Study whose scores dropped from their test at three months to the one at 12 months. The 21 subjects whose scores dropped during this interval of time had a mean Deviation IQ<sup>14</sup> of 123.5 at age 10 years, which is not significantly different from the mean Deviation IQ of the total Berkeley Growth Study group. Neither does low score at a year have any special significance, as can be shown by determining the mean score at ten years of those BGS subjects whose Deviation

<sup>14</sup> The Deviation IQ is defined as the standard score, given a mean of 100 and a standard deviation of 16 (37).

IQ's were below 80 at 12 months of age. The mean Deviation IQ for these four subjects at ten years was 131.7.

If the Hetzer-Wolf tests have no greater predictive value than other infant tests (as such evidence as is available indicates), the conclusion is obvious: Even if these infant's scores were below average we would not need to be concerned, because their scores at one year will be essentially unrelated to their performance by the time they have reached school age. This is, of course, in direct contradiction to the position taken by Spitz, for he maintains that the effects of the maternal deprivation on the Foundling Home children has resulted in irreparable damage (43, 44), and that such damage is reflected in the infants' test scores.

*The Relation of "Deterioration" to Separation in Cases of Anaclitic Depression*

The infants who developed the syndrome of anaclitic depression in the Nursery group were separated from their mothers for a practically unbroken period of three months<sup>18</sup> (45, p. 319). Spitz states that "This removal took place for unavoidable external reasons" (45, p. 319). It is regrettable that he does not specify these reasons so that the infants' subsequent development can be appraised in the light of them. He reports that 45 of the 95 infants (45, p. 336) separated from their mothers show some manifestation of this syndrome (45, p. 318) and that "No child developed the syndrome in question whose mother was not removed" (45, p. 320). According to his

<sup>18</sup> While Spitz maintains that race does "... not appear to exert demonstrable influence on the incidence of the syndrome" (45, p. 318), it should be noted that severe depression occurred among only 9 per cent of the white children whereas it occurred in 26 per cent of the colored children.

graph (45, p. 329), the average age at which mother and child were separated was seven months. The last test before the separation was at six months and one day and the first test afterwards at seven months and twenty-six days. Between these two dates there was a decline in score from 121.5 to 98. Of this 23.5 point decline there was only a 6.5-point drop in score from the first test after separation to the last test before reunion. Subsequent to reunion there was a 25-point rise. Considerations in the next section would appear to raise questions as to whether these changes

TABLE 1

AVERAGE DEVELOPMENTAL QUOTIENTS OF SPITZ'S NURSERY SUBJECTS: A COMPARISON OF AVERAGE DQ'S OF SUBSAMPLES WITH EACH OTHER AND WITH THE TOTAL GROUP\*

Age (in mos.)	Original Group (43, p. 71) N=69	Ana- clitic Depres- sion Group† (45, p. 329) N=45	Re- sidual Sam- ple N=56	Total Group‡ (46, p. 87, 101-103) N=170
2-3	98.0			116.5
3-4	105.0			119.0
4-5	110.0			110.5
5-6	110.0			105.0
6.0	105.0	121.5	90.2	104.5
6-7	100.0			104.0
7-8	110.0			109.5
7.9	110.0	98.0	116.6	109.0
8-10	110.0	91.5	117.3	107.5
10-12	98.0	116.0	82.0	97.5

\* Italicized values were interpolated from Spitz's curves.

† These children were tested at 4 ages: 6 mos. 1 day, 7 mos. 26 days, 9 mos. 15 days, and 11 mos. 0 days.

‡ The total group consisted of 196 cases; however, Spitz's report is based on only 170 of these cases, as the other 26 had not reached the age at which autoerotic activities begin (46, p. 87).

are necessarily the result of separation from and reunion with the mother.

In his article on autoeroticism, Spitz presents a graph apparently based on 170 of the 196 infants thus far observed in Nursery (46, p. 87,



101-103). (The reason for excluding 26 of the cases was that these infants had not yet reached the age at which autoerotic activities usually begin.) Average DQ's for this group of 170 infants were taken from his graph (46, p. 87, 101-103) and are presented here in Table 1. This total group can be divided into three subgroups: (a) 69 infants in the original group. Average DQ's for this subgroup were obtained from the curve in his first article (43, p. 71) and are presented in Table 1. (b) 45 infants separated from their mothers and manifesting anaclitic depression. The graph for this group Spitz presents in his article on anaclitic depression (45, p. 329). (As he did not indicate on how many cases the graph was based or how many were represented at each age, it can only be assumed that it was based on the 45 cases.<sup>16</sup>) Average DQ's for these infants are given in Table 1. (c) Remaining 56 cases of the 170 not accounted for by the preceding two subgroups. From the scores in Table 1 for the total group and for the other two subsamples, the scores for this residual sample of 56 cases were determined for the last six months of the first year and are also included in the table.<sup>17</sup> The fluctua-

tion of the average for these 56 children is marked.

The changes for this residual sample occur at the same ages as for the subjects manifesting anaclitic depression. That is, at the same age that Spitz reports a 29-point drop in score for subjects manifesting anaclitic depression as a result of their being separated from their mothers, this other subsample shows a 27-point rise in score, and at the same age that he reports a 25-point increase in score as a result of the subjects being reunited with their mothers, this other group shows a 35-point drop in score.

As was indicated earlier, it appears that 95 of the children were separated from their mothers and only 45 developed the disorder of anaclitic depression. Spitz does not give the average scores for the other 50 subjects, but if we assume that they are among the 56 subjects of the subsample referred to above, we can determine the average score for the combined group of 101 subjects. The means for the same ages are 104, 108, 106, and 97, respectively. One wonders if the reason for the low scores in the anaclitic depression group is that he assigned to it only those of the 95 cases who showed a marked drop in score. It is impossible to determine from the data he presents whether or not this surmise is correct. However, his contention that the 45 cases manifesting anaclitic depression show a drop and increase in score as a result of separation from and reunion with their mothers would appear to exact little confidence in the light of two other considerations: namely that the other and even larger subsample referred to above shows even more marked fluctuations in

<sup>16</sup> This graph may be based on more than 45 cases as he labels it "Variations of Development Quotient (Average) Under the Influence of Separation from and Reunion with Mother" (45, p. 329), and more than this number were separated from their mothers. (According to Spitz, not all children who were separated from their mothers developed this syndrome [45, p. 320].)

<sup>17</sup> In order to obtain the scores for the residual sample, at each age at which an average score was given for the anaclitic depression group the average score for the total group and for each subgroup was multiplied by the number of cases in the respective samples. At each age the difference between the product for the total sample and the sum of the products for the subsample was divided by the number of subjects in the residual group to obtain its mean score. This procedure

assumes the same number of cases at each age, i.e., it weights the mean value according to the number of cases reported in the sample.

score, and the pattern of means which he gives for the total sample of Nursery is completely out of line with those which McGraw (32) found for unselected samples of white and Negro infants (cf. Fig. 1).

### *Physical Status of the Children*

Spitz gives certain information concerning the physical condition of the infants. With regard to the children in Nursery, he says,

During the *three-and-a-half years*<sup>18</sup> of our study of Nursery, we had occasion to follow 122 infants, each for approximately a full year. During this time *not a single child died*. The institution was visited by no epidemic. Intercurrent sickness was limited, on the whole, to seasonal colds, which in a moderate number developed into mild respiratory involvement; there was comparatively little intestinal disturbance; the most disturbing illness was eczema. The *unusually high*<sup>19</sup> level of health maintained in Nursery impelled us to look into its past record (44, p. 117).

This almost glowing picture of health on 122 infants followed for approximately a full year receives a somewhat different cast when in the second article in this same volume he discusses the presence of anaclitic depression in 45 (45, p. 320) of these infants. He says, "Nor can the infant enact a suicide; but it is striking that these cases *one and all*<sup>20</sup> show a great susceptibility to intercurrent sickness" (45, p. 320). Forty-five children constitute 37 per cent of the total sample. If such great susceptibility to "intercurrent sickness" is shown in 37 per cent of the children, it would seem difficult to maintain without qualification that, on the average, an "unusually high level of health [was] maintained in Nursery."

However, the physical condition of the children of Nursery needs less

attention than that of the Foundling Home children. According to Spitz,

The children in Foundling Home showed all the manifestations of hospitalism, both physical and mental. In spite of the fact that hygiene and precautions against contagion were impeccable, the children showed, from the *third*<sup>21</sup> month on, extreme susceptibility to infection and illness of any kind. . . . No figures could be elicited on general mortality; but during my stay an epidemic of measles swept the institution, with staggeringly high mortality figures, notwithstanding liberal administration of convalescent serum and globulins, as well as excellent hygienic conditions. Of a total of 88 children up to the age of 2½, 23 died. . . .

In view of the damage sustained in all personality sectors of the children during their stay in this institution, we believe it licit to assume that their vitality (whatever that may be), their resistance to disease, was also progressively sapped (43, p. 59).

This quotation can better be appreciated when it is recalled that this was a cross-sectional study at the time this was written, and that his initial study of this group was apparently done in a short period of time, perhaps in less than a week. Thus it appears that the study well may have been carried on at a time when the children were ill. If so, it might be expected that the children would appear to have a lowered vitality. However, one would also think that Spitz would have been surprised at how well the children performed on the Hetzer-Wolf tests, rather than at how "poorly" they did.

Although he notes that from the third month the children showed extreme susceptibility to infection and illness of any kind, he could only assume that their vitality, their resistance to disease, was progressively sapped during their stay in the hospital, as he did not observe the children except for an extremely brief period of time. From this and his

<sup>18</sup> Italics mine.

<sup>19</sup> Italics mine.

<sup>20</sup> Italics mine.

<sup>21</sup> Italics mine.

subsequent articles (44, 45) it is clear that the purported cause is psychosomatic damage resulting from unfavorable environmental conditions. Spitz feels that the most important of these environmental conditions is the separation of the child from his mother. Such a conclusion would hardly seem warranted in regard to health in the fourth and fifth months since "... the separation from the mother took place beginning after the third month, but prevalently in the sixth month"<sup>22</sup> (45, p. 331).

It appears that Spitz's comparisons of these infants before and after separation from the mother warrant little credence, as they were based on cross-sectional data. A less conservative evaluation would maintain that his conclusions must be rejected as he was contrasting two groups of children of whose former and future development he was not cognizant.

To obtain the flavor of Spitz's "longitudinal study" of these infants, the following paragraph is quoted from his article published in 1951; "The progressive deterioration and the increased infection-liability lead in a distressingly high percentage of these children to *marasmus and death*."<sup>23</sup> Of the 91 children followed by us for two years in Foundling Home, 37 per cent died . . ." (47, p. 271). If the reader assumes that 91 children were followed for two years, he is making an erroneous assumption, because 23 of these children (25 per cent) died of measles during the short time that Spitz was initially there (43, p. 59), another four died before the end of the first

year, and 7 died in the second year; hence, it would have been possible to follow only 57 of these children. However, according to his own report, 36 more of these could not be learned about in the follow-up study (44, p. 114), 23 because they had been taken back to their families, 9 because they had been adopted or placed in other institutions, and 4 could not be accounted for. At one point the number followed for two years is clearly stated by Spitz (44, p. 114) to have been 21. It would appear hazardous to base conclusions on such a small portion of the original sample, inasmuch as the selective factors operating in its reduction are unknown.

The preceding quotation from Spitz might lead one to believe that *marasmus* was the most frequent cause of death; however, it should be noted that this "disease" was not even mentioned in his discussion of the death of these 34 infants in either of his first two reports. In the second of these he says, "In the course of the first year, 27 of these died of *various*<sup>24</sup> causes, among which were an epidemic of measles, intercurrent sickness, and cachexia; by the end of the second year, another 7 of those originally seen had died; this represents a total mortality of over 37 per cent in a period of two years" (44, p. 114). As noted above, measles was the most frequent of these various causes of death. It seems surprising that Spitz neglected to mention *marasmus* both in his first report (43) and in his second report (44) since it was present in such a "distressingly high percentage of these children."<sup>25</sup>

<sup>22</sup> In contrast to this group, breast-feeding in the Nursery group was less frequent; thus Spitz says, "In Nursery this percentage is smaller, so that in most cases a formula is soon added, and in many cases weaning takes place early" (43, p. 61).

<sup>23</sup> *Italics mine.*

<sup>24</sup> *Italics mine.*

<sup>25</sup> While in the second article Spitz states that cachexia was one of the causes of death, it should be noted that this disorder was *not* mentioned in the first article on "hospitalism" (43). He does not specify which of the more than a dozen different kinds of cachexia this is,

Regarding these infants' physical development, Spitz gives only cursory information. In his follow-up report on Foundling Home infants, he states that only two have attained normal height of the two-year-old, and that only three of the subjects "fall in the weight range" of normal two-year-olds. This retardation amounts to more than a year in a number of the cases and is indeed very marked; it is especially marked in height, and we are left with the inference that maternal separation is responsible for this retarded growth. It would seem that the prediction may then be made that children separated from their mothers as in Foundling Home will be drastically retarded in growth and height. As we have not usually believed that skeletal growth is to so great a degree responsive to psychological influences, alternative interpretations should be considered: (a) separation from the mother is the cause of extreme retardation in physical growth; (b) this sample (after removal of the 34 cases who died and of the 36 who were not available for the follow-up study) constitutes an inferior biological selection; (c) the norms used by Spitz (not stated) are not applicable to this population (not described).

The first alternative appears to claim more, especially with regard to physical growth, than students of child development are likely to accept without confirming evidence from more carefully controlled studies. The second and third alternatives, apparently discarded by Spitz, must be regarded as reasonable unless negated by further data.

#### SUMMARY

From a study of several groups of infants from contrasting environ-

and as far as the writer has been able to ascertain, no one of the forms is synonymous with marasmus.

ments, Spitz concludes that infants, as a consequence of being separated from their mothers, develop psychological disorders. These disorders are supposedly manifest in both "mental and physical" symptoms. A careful consideration of these studies indicates that they were so planned and carried out that they could give neither positive nor negative evidence for his hypotheses.

Some of the major vulnerabilities in Spitz's report of his studies may be summarized as follows: He fails to indicate the dates and places of the studies and neglects to indicate the composition and training of the research staff. He is inconsistent in his report of the number of children present in his studies, and his descriptions of their parents are contradictory. The groups which are compared as to mental and emotional development apparently differ in racial extraction, socioeconomic background, and in heredity. He is inconsistent in his descriptions of their physical surroundings, of their care, and of their physical health. While most of his conclusions could only be considered plausible if based on longitudinal data, most of them are based on cross-sectional data, or at best on a mixture of the two. His reports regarding the average amount of time spent in observing the children are inconsistent and other considerations indicate that there was a great deal of variability in the amount of time different infants were observed. He uses tests with means which apparently decrease with age. He assumes that test results at one age in infancy are highly related to test results several months later and that the same numerical quotient at different ages has the same relative significance, assumptions which are unwarranted in terms of our present state of knowledge.

The preceding conclusions make it

clear that the results of Spitz's studies cannot be accepted as scientific evidence supporting the hypothesis that institutional infants develop psychological disorders as a result of being separated from their mothers.

#### CONCLUSION

The writer does not doubt the potential advantages of maternal as compared with institutional care. An analysis of institutional care, involving procedures in a large number of institutions, would be expected to show a somewhat different pattern of benefits and of hazards than in the case of children brought up in their own homes. As yet, however, we do not have convincing evidence, based on scientifically controlled investigations, as to any of the major problems in this area.

#### THE FISCHER STUDY OF HOSPITALISM<sup>28</sup>

Fischer (16) comes to conclusions similar to Spitz's in a study of infants in a Catholic home for unmarried mothers, the St. Agnes Home in West Hartford, Connecticut. The subjects used in this study consisted of 62 out of 189 infants tested between the chronological ages of six and seven months in the years 1946 to 1949. The children were tested with the Cattell infant test and those who had IQ's below 90 were chosen for the study.

Fischer maintains that the proportion of children falling below an IQ of 90 exceeds expectancy (62 of 189 children, or 33 per cent). She gives no basis for this contention and it seems highly questionable. If we assume an *SD* of 16 for IQ's on this

scale at the age of six months, we would expect 26.5 per cent of a normal sample of subjects to fall below an IQ score of 90. The difference between this proportion and Fischer's is not significant. (As far as the writer knows, *SD*'s for each month for this scale are not available.)

Fischer considers that the low scores for these infants are the result of their being institutionalized; thus she says, "They most probably are but an indication that the infant in question has reacted early and typically to institutional frustration . . ." (16, p. 531). This contention is obviously unacceptable inasmuch as she chose those children with IQ's below 90 and the number of cases which she found is not significantly greater than one would expect, assuming an *SD* of 16 points. In fact, if on an intelligence test, one did not find a sizable proportion of the cases falling below 5/8 standard deviations at each age, either the test would be suspect or one would suspect a bias in the selection of cases.

The mean IQ for this group of subjects was 76.11. Assuming a chronological age of 6.5 months, this would mean an average mental age of 4.9 months, i.e., the children were on the average 1.6 months retarded. In discussing these children Fischer says,

So far our six-month-old, suffering from what we consider a form of hospitalism, emerges therefore as a baby with good sensory, social and muscular strength reaction who, however, does not engage in the grasping activities which to an essential degree characterize the normal six-month-old (16, p. 528).

From the preceding quotation and description of the subjects, it is apparent that Fischer places a great deal of weight on the child's performance on the Cattell infant scale. It seems almost as if she would maintain that an infant in an institution who makes an IQ score below 90 is

<sup>28</sup> The quotations in this section from L. Fischer, *Amer. J. Orthopsychiat.*, 1952, 22, 522-533, are made with the permission of the American Orthopsychiatric Association, Inc.



afflicted with "hospitalism." However, it is apparent that if one selected children at the age of six months who are retarded to the extent of this sample, that is, children who on the average are performing only at the level of the typical child 4.9 months old, they could not be expected to perform the grasping activities of the average child of six months of age. Hence we cannot accept her conclusion, on the basis of this sample, that lack of this ability is a characteristic of "hospitalism."

Fischer seems to feel that the children subsequently placed in adoptive homes improve in performance on the scale as a consequence of the placement. She states, "Even including these unfavorable scores, the mean IQ of the adopted children (after already increasing in the institution to 86.23) is as high as 97.54. This becomes particularly interesting if we compare the mean scores of all the children in their development" (16, p. 525-526). The mean IQ of the 36 children who were re-examined in their adoptive homes at an average age of 20 months was 97.54. If the tests at these three ages were completely unrelated, one would expect the scores at the subsequent ages to be normally distributed around the mean of the population *which they represent*. However, a slight relationship may be present. For the purposes of this article and for the population which this group represents, the author will assume a mean of 100 at all three ages, while at the same time recognizing that the selective factors in operation make this assumption very vulnerable. If standard deviations were available for the Cattell test for the age levels 6, 11, and 20 months, and if correlations between tests at these age levels were available, scores could be predicted from the initial test at six months to the subsequent

tests at 11 and 20 months. As far as the author knows, such statistical material is not available. Hence, predictions will be made on the basis of material reported in other studies. Continuing to assume an *SD* of 16 at all three ages, and on the basis of Bayley's work (7) assuming a correlation of .52 for tests given at age 6 and 11 months, a correlation of .23 for tests given at 6 and 20 months, and a correlation of .60 for tests given at 11 and 20 months, these scores will be predicted. The predicted scores together with the standard error of estimate are as follows: For 11 months,  $87.6 \pm 13.67$  and for 20 months,  $94.5 \pm 15.57$  and  $91.7 \pm 12.80$ . Granted the previous assumptions, it is apparent that the values obtained by Fischer are not significantly different from those expected purely on the basis of regression toward the mean.

After selecting children retarded in development, children with a mean IQ of 76.1, she states when she comes to a description of their total behavior: "Obviously, the children we observed are youngsters of average potentialities, whose atypical reaction to a test situation is environmentally fostered" (16, p. 529). Since a number of studies have shown that there is very little relationship between tests during infancy and future performance, the test results would not provide a basis for questioning the statement that these subjects "are youngsters of average potentialities." On the other hand the validity of this contention is far from *obvious*, and no information is included in the text of the article which makes it more than plausible. It would seem that such an assumption would have to be based on adequate information about both parents. A negative selective factor is suggested in the "sexual indiscretion" of the mothers and per-

haps by their coming to a maternity home for unmarried mothers. The writer was unable to find in her article any support for the contention that their "atypical reaction" to the test was environmentally fostered.

Fischer subdivides her group of six- to seven-month-old subjects into two other groups. She states,

Again patterning is very clear-cut, as 87 per cent of the children, equally divided, comprise two dominant groups. One of them is essentially passive in all areas—including social and sensory areas. These are children who in the test performance do not react to sound or to the mirror. . . . The other pattern is that of marked social responsiveness—from friendly smiles for the examiner or their own picture in the mirror to tremendous demandingness (16, p. 529).

The two dominant groups would be composed of 23 children each. Fischer states that the first group of 23 children do not react to sound or to the mirror. However, it is difficult to reconcile this statement with her earlier ones: that 94 per cent of the 53 subjects turn to the sound of

a voice, and that "Significant changes, however, occur in all but two items at the five-month level, with an insignificant drop on item 1 (turning to the sound of a bell)" (16, p. 528). Ninety-four per cent of the 53 subjects would include 50 of the infants. If we assume that all "non-reactors" to sound were in this group, 20 of the 23 children would still have reacted to sound.

Fischer does not give a sufficient description of the second group's performance on the test to permit an adequate evaluation; however, from the description given of both groups, the question is raised in the author's mind as to whether the same subgroups could have been obtained merely by dividing the children with IQ's below 90 into two groups, those most and least retarded.

The preceding considerations would seem to raise considerable doubt that Fischer's investigation supports Spitz's conception of hospitalism.

#### REFERENCES

1. ACKERMAN, DOROTHY S. The critical evaluation of the Viennese Tests applied to 200 New York infants six to twelve months old. *Child. Developm.*, 1942, 13, 41-53.
2. ANASTASI, ANNE, & FOLEY, J. P. *Differential psychology*. (Rev. Ed.) New York: Macmillan, 1949.
3. ANDERSON, L. D. The predictive value of infancy tests in relation to intelligence at five years. *Child Developm.*, 1939, 10, 203-212.
4. BAYLEY, NANCY. *The California First-Year Mental Scale*. Berkeley: Univ. of California Press, 1933.
5. BAYLEY, NANCY. Mental growth during the first three years: A developmental study of sixty-one children by repeated tests. *Genet. Psychol. Monogr.*, 1933, 14(1), 1-92.
6. BAYLEY, NANCY. *The California Infant Scale of Motor Development*. Berkeley: Univ. of California Press, 1936.
7. BAYLEY, NANCY. Consistency and variability in the growth of intelligence from birth to eighteen years. *J. genet. Psychol.*, 1949, 75, 165-196.
8. BAYLEY, NANCY, & JONES, H. E. Environmental correlates of mental and motor development; a cumulative study from infancy to six years. *Child Developm.*, 1937, 8, 329-341.
9. BOWLBY, J. Maternal care and mental health. *Bull. World Health Organ.*, 1951, 3, 355-534.
10. BÜHLER, CHARLOTTE, & HETZER, HILDEGARD. *The first year of life*. New York: John Day, 1930.
11. BÜHLER, CHARLOTTE, & HETZER, HILDEGARD (Eds.). *Testing children's development from birth to school age*. New York: Farrar and Rinehart, 1935.
12. CATTELL, PSYCHE. *The measurement of intelligence of infants and young children*. New York: Psychological Corporation, 1947.
13. DENNIS, W. The effect of restricted practice upon the reaching, sitting, and standing of two infants. *J. genet. Psychol.*, 1935, 47, 17-32.

14. DENNIS, W. Developmental theories. In W. Dennis (Ed.), *Current trends in psychological theory*. Pittsburgh: Univer. of Pittsburgh Press, 1951. Pp. 1-20.
15. DENNIS, W., & DENNIS, MARSENA G. Behavioral development in the first year as shown by forty biographies. *Psychol. Rec.*, 1937, 1, 349-361.
16. FISCHER, LISELOTTE K. Hospitalism in six-month-old infants. *Amer. J. Orthopsychiat.*, 1952, 22, 522-533.
17. FRANKL, LISELOTTE, & WOLF, KÄTHE. The test for the first year of life. In Charlotte Bühler and Hildegard Hetzer (Eds.), *Testing children's development from birth to school age*. New York: Farrar and Rinehart, 1935. Pp. 95-127.
18. GELEERD, ELIZABETH R. A contribution to the problem of psychoses in children. *Psychoanal. Stud. Child.* 2, 271-291. New York: International Univer. Press, 1946.
19. GESELL, A. L. *Infancy and human growth*. New York: Macmillan, 1928.
20. Halpern, L. J. *How to raise a healthy baby*. New York: Prentice-Hall, 1940.
21. HARTMANN, H., KRIS, E., & LOEWENSTEIN, R. M. Comments on the formation of psychic structure. *Psychoanal. Stud. Child.* 2, 11-38. New York: International Univer. Press, 1946.
22. HERRING, AMANDA. An experimental study of the reliability of the Bühler Baby Tests. *J. exp. Educ.*, 1937, 6, 147-160.
23. HETZER, HILDEGARD, & WOLF, KÄTHE. Baby tests. *Z. Psychol.*, 1928, 107.
24. HETZER, HILDEGARD, & WOLF, KÄTHE. Baby tests. In Charlotte Bühler and Hildegard Hetzer (Eds.), *The first year of life*. New York: John Day, 1930. Pp. 189-239.
25. HILGARD, ERNEST R. *Introduction to psychology*. New York: Harcourt Brace, 1953.
26. HSU, E. H. On the application of Viennese Scale to Peiping babies. *J. genet. Psychol.*, 1946, 69, 217-220.
27. HUBBARD, RUTH M. A study of the reliability and validity of the Bühler infant scale. *J. genet. Psychol.*, 1935, 47, 361-384.
28. JONES, H. E. Homogamy in intellectual abilities. *Amer. J. Sociol.*, 1929, 35, 369-382.
29. JONES, H. E. The environment and mental development. In L. Carmichael (Ed.), *Manual of child psychology*. (2nd Ed.) New York: Wiley, 1954.
30. KRIS, E. Notes on the development and on some current problems of psychoanalytic child psychology. *Psychoanal. Stud. Child.* 5, 24-46. New York: International Univer. Press, 1950.
31. LOEVINGER, JANE. On the proportional contributions of differences in nature and in nurture to differences in intelligence. *Psychol. Bull.*, 1943, 40, 725-756.
32. MCGRAW, MYRTLE B. Comparative study of a group of Southern White and Negro infants. *Genet. Psychol. Monogr.*, 1931, 10(1), 1-105.
33. MCNEMAR, Q. *The revision of the Stanford-Binet scale, an analysis of the standardization data*. New York: Houghton Mifflin, 1942.
34. MASLOW, A. H., & MITTELMANN, BÉLA. *Principles of abnormal psychology*. (Rev. Ed.) New York: Harper, 1951.
35. ORLANSKY, H. Infant care and personality. *Psychol. Bull.*, 1949, 46, 1-48.
36. PINNEAU, S. R. A critique on the articles by Margaret Ribble. *Child Developm.*, 1950, 21, 203-228.
37. PINNEAU, S. R. Changes in performance on intelligence tests from one month to eighteen years. Paper read at West. Psychol. Ass., Seattle, June, 1953.
38. RIBBLE, MARGARET. *The rights of infants*. New York: Columbia Univer. Press, 1943.
39. RIBBLE, MARGARET A. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 2. New York: Ronald Press, 1944. Pp. 621-651.
40. RUCH, FLOYD L. *Psychology and life*. (3rd Ed.) Chicago: Scott, Foresman, 1948.
41. RUCH, FLOYD L. *Psychology and life*. (4th Ed.) Chicago: Scott, Foresman, 1953.
42. SHIRLEY, MARY M. *The first two years: Vol. I. Postural and locomotor development*. Minneapolis: Univer. of Minnesota Press, 1931.
43. SPITZ, R. A. Hospitalism. An inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal. Stud. Child.* 1, 53-74. New York: International Univer. Press, 1945.
44. SPITZ, R. A. Hospitalism: A follow-up report. *Psychoanal. Stud. Child.* 2, 113-117. New York: International Univer. Press, 1946.
45. SPITZ, R. A. Anaclitic depression. *Psychoanal. Stud. Child.* 2, 313-342. New York: International Univer. Press, 1946.
46. SPITZ, R. A., & WOLF, KATHERINE M.

- Autoerotism. Some empirical findings and hypotheses on three of its manifestations in the first year of life. *Psychoanal. Stud. Child*, 3-4, 85-120. New York: International Univer. Press, 1949.
47. SPITZ, R. A. The psychogenic diseases in infancy: An attempt at their etiologic classification. *Psychoanal. Stud. Child*, 6, 255-275. New York: International Univer. Press, 1951.
48. Terman, L. M., & Merrill, Maud A. *Measuring intelligence*. New York: Houghton Mifflin, 1937.
49. Wolff, G. Maternal and infant mortality in 1944. *U. S. Children's Bur. Stat. Series No. 1*, Fed. Sec. Agency, Soc. Sec. Admin., U. S. Children's Bureau.
50. Milk—and love. *Time*, 1952, 59, (18), 51.

Received July 7, 1954.

## REPLY TO DR. PINNEAU

RENÉ A. SPITZ

*New York City*

Pinneau has expended an extraordinary amount of sagacity and labor in an attempt to discredit every facet of the research I have published. To reply to each of his innumerable points; to correct his misunderstandings, arbitrary conclusions, and unwarranted assumptions would require more space than his article. I will therefore limit my answer to only a few of the most confusing major arguments advanced.

My research, conducted for close to twenty years and comprising the intensive and systematic observation of over 400 infants in their first year, was reported in three or four dozen scientific and professional articles, as well as in about a dozen motion pictures. Of these Pinneau has picked out five articles for the purposes of his polemic. If we disregard the numerous less essential points contained in his paper, we find that he challenges primarily the three following conclusions to which I had arrived in my research:

1. That affective interchange is paramount, not only for the development of emotion itself in infants, but also for the maturation and the development of the child, both physical and behavioral.

2. That this affective interchange is provided by the reciprocity between the mother (or her substitute) and the child.

3. That depriving the child of this interchange is a serious, and in extreme cases, a dangerous handicap for its development in every sector of the personality.

Pinneau endeavors to invalidate these conclusions with the help of statistical manipulations on one hand, and on the other by implying that the experimental psychological material I presented is unsound. To begin with, I wish to make it clear that the experimental psychological and statistical material in the five articles discussed by Pinneau was not introduced by me to *prove* my point, as the articles are addressed to medical readers. They were used as supportive evidence, subordinated to the clinical data—an illustration, as it were, of the description presented. Dismissing these statistics as inadequate would therefore not invalidate my clinical findings. It will, however, be shown further on that the statistics presented by me are correct, and therefore independently confirm and support the clinical findings.

To achieve his purpose, Pinneau has laid a sort of patchwork of numbers. His numerical speculations rest on figures culled, in a biased manner, from five of my publications; these five publications, widely separated in time, deal with different phases of my research. To these mathematics he added various unsupported claims which are in contradiction with the actual facts. With their help he tries to prove that the observations on which my findings and conclusions are based are not longitudinal, but cross-sectional. Once it is proved that they are cross-sectional, he argues, then the statements advanced by me are invalid. Pinneau has not defined what he means by longitudinal. For



the purposes of a study on the first year of life longitudinal has to be defined as comprising a period sufficient to detect significant developmental changes in the subject. In the first year of life such a study will require at least two and preferably three months.

I will begin with Pinneau's treatment of the cases in the original Nursery group. He has confused the figures of the original sample of 69 children observed in Nursery for the purposes of the study on "Hospitalism" (16, 17) with the sample of 123 children observed in the same institution for the purposes of "Anaclitic Depression" (18). He posits an arbitrary date for the beginning of observations in Nursery and thus succeeds in creating the impression that only one-third or at best one-half of the children in question in either sample were observed consistently for a full year. He concludes from this that "the graphs based on the sample of children must to a large extent represent different children at each age," a wording which creates in the reader's mind the impression that most of the children were seen only once or twice in the course of their first year.

I can dispel the confusion created by Pinneau. The 69 children who form the sample of the article on "Hospitalism" (16) consist of:

- 39 children seen for 12 consecutive months or more from birth (within two weeks of delivery);
- 15 children seen for 11 months or more, beginning within the first month of life;
- 9 children seen for 11 months, beginning within the first month of life;
- 6 children seen for 10 months, beginning within the second month of life.

Let me add for those not familiar with infant observation that no test of any kind has yet been devised to provide really useful psychological

information on the infant in its first two months of life, during which behavior is more in the nature of reflex than response. Therefore we did not consider it particularly important that a minority of our sample was not seen by us in the first month and a few of them not in the second month. It might also be mentioned that when Pinneau discusses the usefulness of the Viennese test, most of the discrepancies which he makes so much of are caused just by the unreliability of all and any test scores in the first two months of life. Each of the 69 children was seen at least once a week during the whole period of his stay by myself and/or my associates. It seems to me that the 30 days by which some of the children fall short of the full year's observation hardly can detract from the longitudinal character of the study. When possible each child was tested once a month. When this was not feasible, bimonthly intervals were observed.

After having put in question the original sample of 69 children dealt with in "Hospitalism" (16), Pinneau casts doubts upon the procedures used with subsequent children in the Nursery. The fact is that the populations that were discussed in the studies in my subsequent articles ("Anaclitic Depression" (18), "Autoerotism" (21), "Psychogenic Diseases" (23)) were routinely observed beginning with the first day when the child arrived (that is, within two weeks of birth) at Nursery. A film record was made of its first responses, and the observations continued until the child left the institution, that is, after the completion of the first year.

Pinneau's argumentation on the quantitative aspects is therefore invalid when all the facts and not merely a selected portion of them are examined.

We will now proceed to examine other of Pinneau's treatments of data "to obtain the flavor" (as he puts it) of his criticism.

Perhaps this is the place to mention that a good many of Pinneau's interpretations were made possible by the special circumstances surrounding my research. The physician has a professional obligation of discretion to his patients. Furthermore, penal institutions are forbidden by law to give any clue to the identity of their inmates and the same goes for Foundling Home. I made every effort to keep identity, time, and place protected from uncalled-for inquisitiveness, to the point of using, here and there, misleading clues, such as the term "Western Hemisphere." For the reader's information that term was used by me in a cultural sense, meaning the Western world, including Europe.

When therefore I avoided publishing any specific date on the sample of the 91 children housed in Foundling Home, Pinneau found himself at a loss and therefore had recourse to the following conclusion: "It would seem obvious that in the original study the Foundling Home infants were tested only once, and that the original graphs were based on cross-sectional data." This statement is in contradiction to the facts presented in my studies on "Hospitalism." Pinneau makes this the backbone of his arguments. He variously refers to "the brief" period of time, to "the extremely short time" spent in the observation of Foundling Home, to "the purely cross-sectional nature of the study," and thus he progressively raises his own bid to the point where he has convinced himself that my observation of the children in Foundling Home was limited to "less than a week" and that these children were tested only once.

The facts are as follows: The children in Foundling Home were observed by me for over three months, daily, during four to six hours per day. During this time they were tested, filmed, and observed by myself and my assistants. In the course of the subsequent two years they were observed and photographed at four-month intervals by an assistant trained by myself (17, pp. 114 f., 23, 270 f.).

The technique used in the above two examples of Pinneau's argumentation is applied by him with many variations, modifications, and elaborations. A further example is provided when he indignantly states that the 91 children in Foundling Home were not followed for two years, *because in the course of these two years 34 had died*. He appears not to have grasped a simple medical fact: if the initial sample of a population of 91 is followed for two years, it is the mortality statistics and the condition of the survivors in which the physician is interested. The death of 34 of the original population will provide the physician with the most alarming proof that the conditions for survival are precarious indeed.

The physician will regret that another 36 of the original sample could not be followed because they were adopted or transferred; but he knows that this is one of the shortcomings of longitudinal studies. And the absence of data about these 36 can under no circumstances invalidate the horrifying finding that more than one-third of the original population died within two years. He will realize that there is every probability, even in the best of circumstances, that information about the 36 who successively disappeared from the original sample could only have added to the percentage of the mortality. To state under these circum-

stances that the 21 children who were seen from beginning to end is too small a portion of the original sample for the purpose of concluding that deprivation of maternal care has a harmful effect, appears, to put it mildly, naive.

But what shall we say when we find that Pinneau considers that the 21 survivals constitute "an inferior biological selection"? Does Pinneau really mean to convey that the 37 who died were biologically superior to the 21 who survived?

He raises many other objections which serve to becloud the issue. These appear to me less essential and I will deal briefly with a few of them only.

a. There is an implication that I have attributed the decline of 33 points in the developmental quotient to the consequences of weaning. He supports this allegation through quotations which, taken in their right context, mean just the opposite. On the contrary, I stated in several of my publications that the effect of weaning, particularly around the sixth month, usually results in a rise of the developmental quotient. On the other hand, throughout my publications I stressed that it is the deprivation of maternal contact and of reciprocity with the mother which has destructive consequences. Therefore weaning *as such* is not important; but when children are weaned, they concomitantly lose that contact with their mothers which is implicit in the feeding situation.

b. He questions the value of the Hetzer-Wolf test (3, 8). But contrary to Pinneau's allegations it has been applied to many thousands of infants in many countries, is being used at present, and compares favorably indeed with all other tests (for the first year) known to me.<sup>1</sup> Regard-

ing its predictive value, just as with any other test, a single administration is uninformative and only the trend of a number of successive applications of the test over a period of time is meaningful.

c. He questions the composition of my staff: It consisted of myself, Katherine M. Wolf (one of the two originators of the Hetzer-Wolf test), and of a number of assistants with Ph. D. qualifications trained by Wolf and myself in testing and observing infants.

d. He objects to the method of comparing different institutions with one another, a debatable point. But it is typical of Pinneau's procedure that he neglects to mention that his objection applies only to one of the five studies discussed by him ("Hospitalism" [16]) and that in the remaining studies, groups of infants present at the same time in the same institution, namely in Nursery, were compared with each other.

e. He objects that the question of congenital abnormalities was not touched upon by me and highlights this as one of the uncontrolled variables in the sample. But the nature of the institutions themselves implies that congenital abnormalities were excluded on admission, as the institutions in question were not equipped to deal with them.

f. He alleges that I have not em-

on the Hetzer-Wolf test. She further calls the fact to my attention that the number of children to whom the Viennese test has been applied under controlled conditions can be found in Charlotte Bühler's book: "Kleinkinder Tests" (3) and in the publications which deal with the restandardization of this test in social and cultural groups, e.g., Hofstaetter (9, 10), Maria Wolf (26, 27), etc. My reference to thousands of children was based on the number of children tested in the course of the years of activity of the Viennese University Psychological Institute and elsewhere, without the results of these tests having been published.

<sup>1</sup> Katherine M. Wolf informs me that she does not intend to answer Pinneau's attack

ployed normal groups of infants for purposes of comparison. That idea of Pinneau's is a negative hallucination. Within the publications examined by Pinneau, in "Autoerotism" (21) on page 103, there is a chart filling half the page, and showing the relationship between *normal* infants, infants in Nursery, and infants in Foundling Home—not to speak of the fact that my whole research is based exactly on this kind of comparison.

This refutation of the objections raised by Pinneau could go on and on. I will refrain from going into further details, except to point out again his unfamiliarity with the subject, which becomes embarrassingly evident when he takes me to task because I mention marasmus as a cause of death only in ulterior publications and not in the first ones. He compounds this blunder in a note (footnote 24) where he points to a similar negligence on my part in regard to cachexia. He appears to believe that the two terms serve to designate distinct disease entities. They do not. Marasmus is a symptom, a progressive wasting away, especially, in infants, but also in senescents. The term cachexia also means progressive wasting away when used alone. As an adjective of the "more than a dozen conditions" which Pinneau mentions, it simply means the wasting away occurring in that particular disease. Pinneau takes the secondary effect for the disease.

The lack of clinical orientation is particularly evident whenever Pinneau tries to apply to clinical observation and to clinical methods his statistical standards. He introduces into the observation of the living being an atomistic concept and does not realize that the shortcomings of a test, with random fluctuations of 5 to 15 points are very small matters indeed, when in the follow-up the children in

Foundling Home fall to an average level of 55 points below the "normal" average of 100, and when the essential point which I made is that these emotionally deprived children were dying in a truly horrifying percentage.

Pinneau's yardstick for test validation may be meaningful in evaluating the behavior of mice and rats; certainly his critical procedures will produce no harm there, for we are not responsible for the survival of laboratory animals and their species. The physician deals with human beings and their survival. I do not want to be dramatic or to harrow Pinneau's conscience. But this is not the first attempt by Pinneau to attack clinical fieldwork by purely deductive reasoning. He has tried to invalidate the pioneer work of Margaret Ribble (14, 15) and in the present case he also attacks Liselotte Fischer and Katherine M. Wolf. But his criticism of this line of research implies that we should continue to raise infants in such places as Foundling Home, at the risk of a mortality of 37½ per cent in two years, and wait for a change until we have ironed out the minor fluctuations in the various tests applied. The results of the work of pioneers like Chapin (4), Lowrey (11), Goldfarb (6, 7) and Bakwin (1) has fundamentally changed our approach to the raising of infants in the last 40 years. The results achieved by the later workers, Anna Freud (5), John Bowlby (2), Margaret Ribble (14, 15) and I believe I can include myself, has been by this time applied all over the world in the practice of many hundreds of hospitals and institutions, with a concomitantly demonstrable saving of innumerable human lives. The exertions of Pinneau will not stop the progress in the care and the understanding of infants achieved in the last 40 years. When faced with



this kind of criticism, we can say with Henry Poincaré: "On ne discute pas avec un physicien, on répète ses expériences!" The quality of Pinneau's criticism has been aptly characterized in Joseph Stone's presidential address to the N.Y. State Psychological Association. There, in a few masterful paragraphs of comment on Pinneau's criticism of Margaret Ribble's work, Stone includes the remark: "I commend you his article [Pinneau's] as a kind of hydrogen bomb perfection of destructive criticism; not a paragraph is left standing for miles around" (24).

Nobody realized better than I the shortcomings of my experimental design, and those of my publications. They are largely due to the exiguity of the means at my disposal, to the space restriction imposed on such

publications in psychiatric journals, and, last but not least, to the fact that I was exploring uncharted territory in which progress can only be halting. But Pinneau, after all, could have asked me to explain apparent inconsistencies and to fill out gaps, if it was information and the advancement of science he was after. He chose not to do this.

Criticizing, questioning, discussing empirical and nonempirical studies in science increases our insight, broadens our knowledge and is highly welcome when based on factual data. A critical discussion, however, built on inference and implication, and which has recourse to invention, serves no useful purpose. On the contrary; it beclouds the issue, it confuses and, moreover, holds up unnecessarily the progress of science.

#### REFERENCES

1. BAKWIN, H. Loneliness in infants. *Amer. J. Dis. Child*, 1942, 63, 30-40.
2. BOLWBY, J. Maternal care and mental health. *World Health Organ. Monogr. Series*, 1951, No. 2.
3. BUEHLER, CHARLOTTE, & HETZER, HILDEGARD. *Kleinkinder Tests*. Leipzig. Barth, 1932.
4. CHAPIN, H. D. A plea for accurate statistics in infants' institutions. *Arch. Pediatr.*, 1915, 32, 724-726.
5. FREUD, ANNA, & BURLINGHAM, D. *Infants without families*. New York: International Univer. Press, 1944.
6. GOLDFARB, W. Effects of early institutional care on adolescent personality. *J. exp. Educ.*, 1943, 12, 106-129.
7. GOLDFARB, W. Effects of early institutional care on adolescent personality. Rorschach data. *Amer. J. Orthopsychiat.*, 1944, 14, 441-447.
8. HETZER, HILDEGARD, & WOLF, KATHERINE. Baby tests. *Z. Psychol.*, 1928, 107, 62-104.
9. HOFSTAETTER, P. R. Testuntersuchungen an japanischen Kindern und das Reifungsproblem. *Z. Kinderforsch.*, 1937, 46, 71-112.
10. HOFSTAETTER, P. R. Was besagen Testergebnisse. Ein Beitrag zum Dimensionsproblem des Entwicklungstests. *Z. Kinderforsch.*, 1938, 47, 92-96.
11. LOWREY, L. G. Personality distortion and early infant care. *Amer. J. Orthopsychiat.*, 1940, 10, 576-585.
12. PINNEAU, S. A critique on the articles by Margaret Ribble. *Child Developm.*, 1950, 21, 203-228.
13. PINNEAU, S. The infantile disorders of hospitalism and anaclitic depression. *Psychol. Bull.*, 1955, 52, 429-452.
14. RIBBLE, MARGARET A. *The rights of infants: Early psychological needs and their satisfaction*. New York: Columbia Univer. Press, 1943.
15. RIBBLE, MARGARET A. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. II. New York: Ronald Press, 1944.
16. SPITZ, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal. Stud. Child*, 1, 53-74. New York: International Univer. Press, 1945.
17. SPITZ, R. A. Hospitalism: A follow-up report on the investigation described in Vol. I. *Psychoanal. Stud. Child*, 2, 113-117. New York: International Univer. Press, 1946.
18. SPITZ, R. A., & WOLF, KATHERINE M. Anaclitic depression. *Psychoanal. Stud. Child*, 2, 313-342. New York: International Univer. Press, 1946.



19. SPITZ, R. A. Environment vs. race. *Arch. Neurol. Psychol.*, 57, 1947.
20. SPITZ, R. A. The importance of mother-child relationship during the first year of life. *Ment. Health Today*. Wash. Soc. Mental Health, 1948, 7, 7-13.
21. SPITZ, R. A., and WOLF, KATHERINE M. Autoerotism. *Psychoanal. Stud. Child*, 3/4, 85-120. New York: International Univer. Press, 1949.
22. SPITZ, R. A. Three first steps in growing up. *Child Study*, 1950/51, 28, 2-5.
23. SPITZ, R. A. The psychogenic diseases in infancy. *Psychoanal. Stud. Child*, 6, 255-275. New York: International Univer. Press, 1951.
24. STONE, J. A critique of studies of infant isolation. *Child Develpm.*, 1954, 25, 9-20.
25. WOLF, KATHERINE M. Observation of individual tendencies in the first year of life. In M. J. E. Senn (Ed.), *Problems of infancy in childhood*. New York: Josiah Macy Jr. Foundation, 1953.
26. WOLF, MARIA. Kleinkinder tests. *Arch. ges. Psychol.*, 1935, 94, 215-246.
27. WOLF, MARIA. Kleinkinder tests an Wohlstandskindern. *Z. Kinderforsch.*, 1935, 44, 191-193.

REPLY TO DR. SPITZ<sup>1</sup>

SAMUEL R. PINNEAU

*University of California*

In replying to my evaluation of his research on infantile disorders, Spitz presents a number of statements which may be examined in relation to specific sections of the original review.

## THE RESEARCH REPORTS REVIEWED

Spitz considers that in dealing with only five articles, I have used an incomplete sample of "three or four dozen scientific and professional articles" and "about a dozen motion pictures."

A search of the *Psychological Abstracts*, the *Quarterly Cumulative Index Medicus*, and the *Biological Abstracts* disclosed from 1935 to the present, 23 articles by Spitz and four films which report aspects of his research on infants.

My choice of articles was dictated by four considerations: Of the articles I read, these gave the most detailed consideration of the institutions, the infants, and their background. All were available in the same Annual (13). In American psychology, these

are the most widely cited among his publications. In his references, he himself refers to them more frequently than to other reports.

## THE STUDIES

In his reply, Spitz has failed to give essential information about the nature of his studies. The Nursery study was apparently initiated in 1942 and that of Foundling Home in 1944 (4, p. 432), at which times his residence is listed as New York (12); data are still withheld concerning the social and geographic areas served by these institutions. The reply suggests additional questions concerning the comparability of the samples, inasmuch as Nursery is apparently in a penal institution in New York (4, p. 433), while the location of Foundling Home is no longer limited to the "Western Hemisphere," but rather to the "Western world." While a physician's responsibility to his patients or subjects is not to be denied, it is difficult to believe that in a matter of this sort professional ethics or legal restrictions could be violated by identifying the institutions concerned, or at any rate by giving de-

<sup>1</sup> The writer wishes to acknowledge the criticisms and suggestions of Harold E. Jones in the writing of this reply.

tails as to national, educational, and socioeconomic samplings involved.

#### THE SAMPLES

Spitz states in his first article (5, p. 56) that 130 infants constituted the *total* population of both institutions, of which 69 were in Nursery and 61 in Foundling Home. Elsewhere in the article the number in Foundling Home is stated as 88 (5, p. 59) and in a different article, 91 (9, p. 271). In the reply he refers to 91 infants in this institution, and states that "If the initial sample of a population of 91 is followed for two years, it is the mortality statistics and the condition of the survivors in which the physician is interested." Very true; but the fact remains that the number of cases in this institution actually followed for two years was 21 and not 91, unless all those that have died are automatically registered as members of subsequent follow-ups.

#### THE NATURE OF THE OBSERVATIONS

In his reply, Spitz states that his original study of the Foundling Home infants required three months. This clarification is important since his earlier statements, on which I based my estimate of the length of time required for the original study (4, p. 435), indicate repeated tests on these children over a two-year period (cf. 4, p. 438). (The graph for this group is reproduced in Fig. 1 (4, p. 440)). While it must be granted that three months of observation can be considered a sufficiently long period of time to qualify the study as a longitudinal one for *some* purposes, it is quite obvious that it cannot be used to show the development of a constant number of Foundling Home subjects during their first year of life. Spitz still does not specify the number of children observed at each age,

nor the number of times the individual children were tested. This may seem to be a trivial matter, but lack of information in such areas and repeated discrepancies, such as noted above, make it difficult to evaluate both the reports of the observations and the interpretations which are offered.

With regard to the Nursery infants, we are informed that the total sample in the institution was included (5, p. 56), that observations began in the first two months, and that each was observed for a period of 12 to 18 months (7, p. 313). Thus, at the time the study began on the original sample of 69 cases, there could have been no children in the institution beyond two months of age. But we know that Nursery had been in existence for at least ten years (6), that it admitted children at frequent intervals (4, p. 433), and kept them through their first year (5, 6). A check on the earlier records of the institution might explain this apparent discrepancy, if the institution could be identified.

#### THE RESULTS

*Clinical findings.* As previously noted in detail, Spitz's clinical findings appear to me to contain numerous contradictions, (see, for example, 4, p. 434, 437, 445, 446). In addition, the statements concerning the care of the children in the two institutions are contradictory (cf. 4, p. 437-438), there are known differences between the two groups that could at least in part account for the results obtained (cf. 4, p. 433-435, 447), and there is lack of information on a number of other variables which might account for the results (cf. 4, p. 437, 447).

He does not deal with any of these specific points in his reply except that of congenital abnormalities. With

regard to these he states that the nature of the institutions implies that they were excluded on admission; however, this statement is difficult to reconcile with his description of the Foundling Home infants' medical care (cf. 4, p. 437-438; 5, p. 62) and his statement that in advanced extreme cases the picture of these children varied from stuporous deteriorated catatonia to agitated idiocy (7, p. 331).

In his articles Spitz lists marasmus and cachexia among the causes of death of the Foundling Home children; however, in the reply he also maintains that these are just symptoms. The question arises as to whether these symptoms, as he now refers to them, were the cause of the death or if the cause was a "disease entity" of which marasmus was the symptom. Measles was listed as the cause of 65 per cent of the deaths; the particular disorders leading to death in the other 35 per cent were unspecified. It should be noted that the distinction between "symptoms" and "disease entities" is not always a clear one, and that apparently not all investigators in the medical sciences would agree with Spitz regarding marasmus (1, 2, 11; cf. 3, p. 221-222 for a summary of these papers).

*Statistical findings.* In his studies Spitz has used the Hetzer-Wolf test. An evaluation of the test in terms of published research indicates that it was inadequately standardized and that a given DQ does not have the same relative significance at different ages within the first year. The available evidence does not support Spitz's contention in his reply that the predictive value of this test is increased by having repeated measurements on the same children during their first year. While it is undoubtedly true that the test has been and is used extensively in many countries, fre-

quency of use cannot, in and of itself, increase the validity of the test.

According to Spitz, the most prominent feature of hospitalism, manifested by the Foundling Home infants, was severe developmental retardation as determined by the Hetzer-Wolf test (5). In his reply he continues to maintain that the drop in DQ score is to be accounted for by the separation of the child from his mother. The average DQ of the children fell from approximately 131 to 72 in the first year; however, the drop from approximately 131 to 76 took place before the prevalent age at which the children were separated from their mothers (cf. Fig. 1 [4, p. 440]). In his first article (5) he states that by the end of the second year the DQ sinks to 45, "... an average level of 55 points below the 'normal' average of 100..." (10). It is to be noted, however, that the individual children on whom this statement was based were not tested over a two-year period, but only over a three-month period, and the DQ decline as reported must therefore involve different samples or different combinations of cases at different ages, the amount of repeated testing and of sample overlap not being specified.

The total Nursery sample of 196 cases was divided into three subsamples in my original article: the original group, the anaclitic depression group, and a residual sample. It was suggested that Spitz chose those subjects showing a marked drop in score for the anaclitic depression sample. Until he has specified that this is incorrect, and why, it can only be assumed that the results in this study can be accounted for in this manner.

#### SPITZ'S CONCLUSIONS

Spitz lists three conclusions in his

reply which he feels that I have challenged. I have, however, not challenged these as propositions, but have merely raised questions as to whether they are in fact supported by the evidence which he presents. He says that my evaluation of his research "... implies that we should continue to raise infants in such places as Foundling Home..." (10). This is incorrect. The purpose of a critical evaluation of a research report is to determine if the data reported warrant the conclusions drawn by the author. I do not believe that Spitz's data do.

#### CONCLUSION

Dr. Spitz's articles are almost the

only research reports in the first six volumes of *The Psychoanalytic Study of the Child* (13) which attempt a statistical analysis of the data. He is to be commended for this; certainly in this area of clinical hunches and hypotheses we have special need for more careful scientific investigations and checking of hypotheses, and the use of statistical materials may be a step in this direction. It may well be that the burden of blame for the uncritical acceptance of his works does not rest with Spitz, who has published his results as he sees them, but rather with those who have acclaimed his work, and whose research training should enable them to make a critical evaluation of such research reports.

#### REFERENCES

1. HISLOP, W. A. Treatment of marasmus by injection of an extract of adrenal cortex. *Lancet*, 1938, **235**, 308-310.
2. HOLT, L. E. *The diseases of infancy and childhood*. New York: Appleton, 1918.
3. PINNEAU, S. R. A critique on the articles by Margaret Ribble. *Child Developm.*, 1950, **21**, 203-228.
4. PINNEAU, S. R. The infantile disorders of hospitalism and anaclitic depression. *Psychol. Bull.*, 1955, **52**, 429-452.
5. SPITZ, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal. Stud. Child*, **1**, 53-74. New York: International Univer. Press, 1945.
6. SPITZ, R. A. Hospitalism: A follow-up report. *Psychoanal. Stud. Child*, **2**, 113-117. New York: International Univer. Press, 1946.
7. SPITZ, R. A. Anaclitic depression. *Psychoanal. Stud. Child*, **2**, 313-342. New York: International Univer. Press, 1946.
8. SPITZ, R. A. and WOLF, KATHERINE M' Autoerotism. Some empirical findings and hypotheses on three of its manifestations in the first year of life. *Psychoanal. Stud. Child*, **3/4**, 85-120. New York: International Univer. Press, 1949.
9. SPITZ, R. A. The psychogenic diseases in infancy: An attempt at their etiologic classification. *Psychoanal. Stud. Child*, **6**, 255-275. New York: International Univer. Press, 1951.
10. SPITZ, R. A. Reply to Dr. Pinneau. *Psychol. Bull.*, 1955, **52**, 453-459.
11. ANONYMOUS. Infantile marasmus. Current comment. *J. Amer. Med. Ass.*, 1930, **95**, 1, 538-539.
12. *American Medical Directory*. (17th and 18th Ed.) Chicago: American Medical Association, 1942, 1950.
13. *The psychoanalytic study of the child*, 1-4. New York: International Univer. Press, 1945-1951.

## BOOK REVIEWS

GOOD, CARTER V., & SCATES, DOUGLAS E. *Methods of research*. New York, Appleton-Century-Crofts, 1954. Pp. xx+920. \$6.00.

This is a lengthy volume addressed to "field workers, graduate students and members of the senior division of the undergraduate college who would evaluate the quality of conclusions, either as producers or consumers of research." It carries the subtitle: "Educational, Psychological, Sociological," an indication of the authors' aim to furnish a treatment of certain research methods common to education and the social sciences. Although the main effort is addressed to students in education, the subtitle is justified in a limited degree, both by the kinds of methods which are discussed and by the use of illustrative materials drawn from psychology and sociology.

Two preliminary chapters, one entitled "Research as a Way of Progress," and the other, "Formulation and Development of the Problem; Research Programs and Needs," introduce the main subjects. These are: literature and library techniques; the historical method; the descriptive method (considered from the points of view of analysis, classification and normative research); the experimental method; methods used in case and clinical studies, and in genetic and developmental studies; and the reporting and implementation of research.

From the point of view of the needs of students of psychology, whether undergraduate majors, graduate students, or established scholars in the field, the values in the book are somewhat uneven, and in a certain sense collateral. The first chapter on "Research as a Way of Progress," for

example, is uncritically inspirational, being based on the thesis that "such strength as will be demanded for survival and for the good life can come only through research; not primarily research in the physical sciences—granted that such research must be continued on a large scale and probably made more fundamental—but research on all of the many facets of living." Now this is an unobjectionable sentence, if an uncritical view is taken of the meaning of the word *research*. However, if by research is meant methodology, particularly methodology as it is defined by the content of this book, the sentence is debatable since it leaves unmentioned the influence of the theorist and, to pass to another universe of discourse, those forces in the development of knowledge which seem to function like the forces closing a gestalt—"discoveries" have been made, it has often been noted, because the time was ripe. In such cases, the men were less important than the ideas with which they were working.

The main treatment suffers from a certain dilemma in which the authors are caught. They do not, they indicate, want to write a textbook on techniques—statistical, psychometric, or sociometric—and neither have they defined their audience as one which can be expected to be competent in the use of such techniques. On the other hand, they are obliged for various reasons to come to terms with all the ways in which problems of research are problems in statistics, in experimental design, or in validation—to give but a few illustrations. The results are not always happy. Their aim is accomplished in too many instances, of importance to the



training of psychologists, by mere mention of the names of techniques and methods—sometimes in lists which convey nothing but the name—by referring the student to other sources, or by passing judgment upon the applicability of techniques in an authoritative way. (No serious student should be left depending upon a professor's authority when it comes to such a matter as the applicability of a technique.) Too seldom is an account offered of their use which illuminates the problems for which they are suitable, or gives detailed guidance in their use, or alerts the student to the special difficulties which must be overcome in their application, or most important of all, leaves the student with a firm grasp of the internal logic of the method which fits him to be his own authority.

The work is strongest where it is sensitive to needs which are either shared by all three kinds of students to whom it is addressed (in education, psychology, and sociology) or to the needs of the students of education alone. Thus it contains a useful treatment of the methods of the historian—about which psychologists ought generally to know more—with practical guides to historical sources, to note-taking methods, to the very important techniques of criticism, both external and internal which are the special virtue of the professional historian, and to the problems of historiography. It contains a useful, but brief, account of questionnaire and interview techniques, and of the problems encountered in the conduct of the kind of educational survey with which professional educationists are familiar. More superficial, but still useful, are the accounts of content analysis, small-group study, and the conduct of research in association with casework.

The book has a number of collateral values which make it a useful reference work. First are the extensive bibliographies. These are organized as footnotes (of which there are nearly 1,600, many of which contain dozens of individual citations), and as Selected References (of which there are nearly 2,100). It is quite possible that the book contains more than 6,000 citations in all. Further, the various bibliographies abound with references as recent as 1954. Second are the many side references, principally of an historical turn, which are a tribute to the wide-range interest of the authors and which enrich the reader—we learn of Hull's practice of keeping a notebook in his graduate days at Wisconsin, that at the end of his career the notebooks totaled twenty-seven volumes, and that some of his stimulus to systematic thinking came from these notebooks, and similarly of Ranke, Mommsen, Huxley, Boring, Tolman, Hall, Bingham, Niebuhr, Sumner, Giddings, Ehrlich, Gibbon, and Kettering, who are merely a few among many others.

Despite these values, however, the work carries implications which will trouble many readers. It implies the primacy of data collection and treatment over the process of reflection from whence come the theories and hypotheses which direct the choice of data to be collected. It carries implications that all kinds of data collection are equally respectable, from an intellectual point of view, as indeed they are where *only* procedural questions arise. But this latter assumption blurs important distinctions in the minds of students. The engineers know better than this when they distinguish explicitly between research and development. It implies that the student can be trained to "evaluate the quality of

conclusions" on the basis of acquaintanceship and reference knowledge, i.e., knowledge of where to go in order to acquire the competency which will make one capable of asking the significant questions in the evaluation of the quality of conclusions. All of this seems to the reviewer to be likely to produce bystanders rather than participants in scientific work.

MALCOLM G. PRESTON

*University of Pennsylvania*

BRAND, HOWARD. (Ed.) *The study of personality: a book of readings*. New York: Wiley, 1954. Pp. xvi + 581. \$6.00.

*The Study of Personality*, as its subtitle accurately reflects, is a collection of readings, consisting mainly of papers that have previously appeared in journals. The editor has contributed four original chapters, including a general introduction and an introduction to each of the three sections into which the book is subdivided; he has also contributed a brief commentary preceding each of the reprinted papers.

The three sections of the book deal, respectively, with theory, methods, and problems. For inclusion in each of the sections the editor has selected material "from clinical, experimental, and social psychology, from anthropology, and from sociology." In his preface, he has indicted his awareness that such broad coverage has been achieved at the expense of omitting many important papers in the field of personality. However, as his avowed purpose was to "show the student how much variety there is in research activity touching on personality," the editor has not committed himself to any selection criteria in terms of which his choices can be criticized. One can only comment, in this connection, that he has

selected an impressive collection of high-level treatments of personality, which have been written from many divergent points of view. The list of contributors reads like a condensed "Who's Who in the Social Sciences."

The editor's emphasis is mainly methodological, and even the section dealing with theory stresses principles of theory construction rather than attempts to formulate a single logically consistent theory of personality. This methodological emphasis, in the reviewer's opinion, definitely restricts the "student" group for which the book can serve a useful purpose; the group of potential readers is probably limited to graduate students and social scientists who have completed their formal training. Graduate students should be able to utilize it as a nucleus around which to organize their reading of significant research contributions to personality theory. Fully trained social scientists will likely appreciate its value in providing cues to the recall of significant contributions that have undergone a process of partial forgetting.

Although the editor does not explicitly state that his book is expected to be of use in the teaching of undergraduates, he does not disclaim its usefulness for this purpose. On this point, the reviewer is inclined to believe that, while the book may become a rich source of lecture material, it cannot be understood adequately by the average undergraduate reader because of its level of sophistication, its level of difficulty, and its specialized emphasis.

The clinical psychologist, and the instructor with clinical inclinations, will likely find this collection of readings somewhat disappointing. There is little in it to satisfy his ideographic propensities, since the attitude of the

editor appears to have been that of assuming that the scientific method in psychology requires a nomothetic approach. That ideographic approaches can be just as "scientific" as any of the nomothetic approaches has been amply demonstrated in the psychological literature. The reviewer, at any rate, is in full agreement with the clinical psychologists and psychiatrists who insist that the investigation of a single personality must be a process of entertaining, testing, and confirming, modifying, or rejecting hypotheses which develop out of the data as they are being collected, whether the data consist of responses to diagnostic tests or responses occurring during psychotherapeutic interactions.

In summary, this collection of readings should fill the need for supplementary reading in a graduate level course dealing with nomothetic approaches to the investigation of personality, or as a supplement to a course in the theory of personality when either the instructor or a good textbook contributes a single consistent theoretical point of view. Psychologists and other social scientists will find it useful as a review of significant and more or less familiar material that should not be allowed to become lost in seldom consulted back issues of the journals.

BERT R. SAPPENFIELD

*Utah State Hospital*

KERMAN, EDWARD F. *What is electroshock therapy?* New York: Exposition Press, 1954. Pp. 152. \$3.50.

Although biased in favor of electroshock, this book should serve a useful purpose in supplying nontechnical authoritative answers to questions usually asked about this popular form of treatment by patients and their relatives.

JAMES D. PAGE

*Temple University*

RYAN, THOMAS, & SMITH, PATRICIA. *Principles of industrial psychology.* New York: Ronald, 1954. Pp. xiv+534. \$5.50.

According to Ryan and Smith, "the book is designed as an introductory survey of the entire field of industrial psychology." They are "concerned with formulating for the reader the principles of industrial psychology."

Since the authors omit many topics, this book cannot be considered a survey of the entire field. In addition, the reader may have some difficulty in identifying the principles from the textual material and so it appears as if this is more a title of a book rather than a clearly stated and outstanding series of principles that lead to a systematic presentation of industrial psychology.

The book presents in substantial fashion those aspects of industrial psychology as the authors perceive it to be. While one may disagree with their point of view, it is clear that the authors are very familiar with the subject matter.

The style of presentation is characterized by critically evaluating research studies reported in the literature and emphasizing the necessary statistical concepts and techniques related to selection of employees. The heavy statistical involvement may make this book a little too difficult for the typical undergraduate student who is not a psychology or statistics major. Admittedly, statistics is an important tool of the industrial psychologist, but whether statistics and industrial psychology can be learned simultaneously from the same text is questionable. One may also wonder whether the student will become confused as a result of the authors' criticisms of other researchers' work. Many references seem to be included

to illustrate shortcomings of research in the field. This is contrary to the usual practice in basic texts of referring only to the best work done in the field and emphasizing the positive aspects of the findings. A completely interested (possibly graduate) student will benefit considerably from such a presentation. The ordinary ones may become annoyed.

Part I, "Selection and Placement," is a solid and conservative presentation of tests and other selection procedures with a considerable emphasis on statistics. Slightly more than 50 per cent of the entire book is devoted to selection and placement. Part II is devoted to "Factors in Efficiency." The first sentence states that "All of this book is concerned with the ways in which efficiency of the performance of workers can be improved." Ryan and Smith apparently view this as the main task of industrial psychology. In addition to considering the concept of efficiency, such environmental factors as lighting, ventilation, noise and their effect on efficiency as well as a critical chapter on time and motion study are included.

There are three chapters on industrial motivation in this book. The authors reject what they term the behavioristic and the Freudian biases and propose a general theory on the initiation and control of activity. They state "that man behaves with respect to the future and past as they anticipate or remember them." According to Ryan and Smith, "the first stage in explaining behavior is a problem of perception, remembering, imagining, understanding and thinking."

From this take-off point, a series of thirteen hypotheses are postulated in the second chapter on motivation. Two examples are: Hypothesis I—Every individual wishes to believe

or perceive himself as valued by the social group as an improved and useful member; Hypothesis IX—Developing skill in any activity tends to enhance the goal character of the activity.

The third chapter on motivation covers such topics as job satisfaction, attitudes, morale. Brief references are made to the Hawthorne studies, the work at the University of Michigan, and experiments done by some of our English colleagues. In addition, boredom, monotony and restriction of output are mentioned. Additional chapters on facilitating learning and accident control conclude the authors' coverage.

Ryan and Smith view the contributions of industrial psychology as (a) placement, (b) evaluating efficiency, (c) motivation, (d) training methods, and (e) accident proneness. If one is content with viewing industrial psychology as the above-mentioned areas and if one prefers to have slightly more than 50 per cent of the text on placement, then this text will be very satisfactory. There is a sound and critical presentation of the material covered. The major novelty of the text is the presentation of a theory of motivation. Whether it "catches" remains to be seen. From a reviewer's point of view it seems to be a description of the ways people might behave rather than a systematic presentation of causality. It does have shades of Woodworth, Gestalt, and Tolman and quite a point is made in rejecting Watson and Freud.

*Principles of Industrial Psychology* is an interesting book for a sophisticated audience. It may be misunderstood by typical undergraduates and it may not be too appealing to the man in industry who wishes to apply some principles.

MILTON L. BLUM  
*College of the City of New York*

- OESER, O. A., & HAMMOND, S. B. (Eds.) *Social structure and personality in a city*. New York: Macmillan, 1954. Pp. xxii+344. \$4.50.
- OESER, O. A., & EMERY, F. E. *Social structure and personality in a rural community*. New York: Macmillan, 1954. Pp. xiii+279. \$3.75.

This two-volume overview of the sociological and psychological structure of the Australian Commonwealth, its people and its institutions, is admirably unique in many respects. The reported research was initiated by the Faculty of Psychology at the University of Melbourne, and aided by local and Unesco grants for the study of social tensions. It is important to note, however, that the work was carried on not only to render contributions to scientific knowledge, but to serve also as instructional devices for University classes in Collective Behavior (Social Psychology). Quite apart from the value of the material presented, it stands as a tribute to a style of cooperative student participation in professional research and reporting that would be unusual anywhere. Considering the reported personal autonomy of Australians, it is the more remarkable there.

It is not uncommon for single research programs to be oriented to the perceptual framework of the respondents used in the study. Indeed, this is quite characteristic of field studies in sociology and cultural anthropology. It is uncommon, however, to find this frame of reference saturating the psychological exploration of a total national culture as is done in this project. The reader is, for example, spared a dreary breakdown of class and caste distinctions as deduced by the investigators. Instead he is exposed to the self-revelations of the population with respect to their perceived roles and their role-

tied assessment of institutions and social strata. Early in the first volume the reader is prepared for this more functional method of interpretation by a so-called "life phase diagram." The diagram schematically emphasizes individual growth, not as a function of chronological age, but as the phase-like passage through institutional groups. The adjustment of the individual to these social settings, his interpretation of them, and his evaluations of their interrelationships with him form the nucleus of the research methodology. The advantage gained by this approach is quickly apparent in its highly objective, yet never coldly detached, portrayal of Australian "culture." One looks in vain for symptoms of creeping national pride, or its reverse—the sterility of "guinea-pig" reporting. Neither is grossly apparent, and the native Australian would quickly identify with the detailed portrait of himself drawn by the authors.

There are many surprises for the American reader who "knows" Australians as a result of war-time experiences, single contacts with visitors, or Sunday supplement stereotypes. He will discover the Australian to be quite different from himself in expressions of nationalism, in "democratic" ideology, in cooperativeness, in neighborliness, and in a host of specific social behaviors. This reviewer would confidently recommend Volume I as a psychological Baedeker for the American traveler in urban Australia. Volume II may be no less authoritative as a rural guide, but it appears to be somewhat more academic and "reasoned" in its treatment of the material.

Without prejudice to the other contributors, special mention should be made of the extraordinarily fine development of the material by S. B. Hammond, and the chapters in which



P. G. Herbst details his conceptualization of family structure. It is Hammond who gives the initial orientation to the inquiry, and who acquaints the reader with the necessary background of history, existent social institutions, and the general way of life. At best this is a tricky job—and at worst is apt to be superficial. Hammond has avoided the attendant dangers and, in a lucid unfolding of the theme, paves the way with both general observations and specific reference to research data. In addition, his interpretive summaries seem to jell what might otherwise be scattered and sparse information.

Herbst's contribution is of another sort. Using a topological description of family structure, he has provided—at least within the framework of the Australian culture—a conceptual tool for characterizing the family both as a field of activity regions (through which family members pass developmentally) and as a locus of interaction patterns for family members. Family tension is dynamically linked with family structure. This is no esoteric abstraction, as handled by Herbst, but an experimentally implemented system which holds promise of a definitive comparison of cultures.

Despite these many attributes, the volumes do have certain shortcomings which detract somewhat from their readability. There is apparent in them a degree of disorganization and asymmetry that is at times disconcerting. One gets the impression that although the general editor did yeoman duty in connecting successive reports, the theme development and writing style of individual contributors did not always lend themselves to smooth internal blending. Volume II, dealing with rural community patterns, is better organized,

but fails to parallel as well as it might the techniques and types of interpretation found in the volume on urban structure.

What appears to be another defect—instances of limited samplings of a total population of over eight million people—is actually not as damaging as it first appears. The studies of Hammond and Herbst, for example, show an extremely high consistency of behavior patterns apparent in family relations. Participation patterns of husband and wife are predictable with 92 and 95 per cent accuracy, respectively. In a culture where behavior is sufficiently homogeneous—as appears to be the case here—the sample size becomes less crucial.

Finally, there is an inconsistent policy of statistical reporting. Throughout Volume II there are adequate indices of significance appended to most tables. In Volume I, however, one must look long and hard for any such criteria.

In no sense do these limitations seriously lessen the worth of the books for a rather wide assortment of potential readers. For the psychologist, sociologist, anthropologist, and political scientist, they should constitute an invaluable reference book, both for content and methodology, in the delineation of national mores and their psychological origins.

DAN L. ADLER

*San Francisco State College*

BURT, CYRIL. *The causes and treatment of backwardness.* (Rev. Ed.) New York: Philosophical Library, 1953. Pp. 128. \$3.75.

This book is a revision and expansion of earlier lectures and writings of Dr. Burt. "The subnormal child has already formed the subject of numerous inquiries and researches; and it seemed to me that the most useful thing that a psychologist could do

would be to survey what is already known and what has already been accomplished, and then to summarize the main conclusions in a form available for the ordinary teacher." Although Dr. Burt's survey is fairly incomplete, it could be read with some profit by the grade school teacher.

SEYMOUR B. SARASON

*Yale University*

KUGELMASS, I. NEWTON. *The management of mental deficiency in children*. New York: Grune & Stratton, 1954. Pp. xii+312. \$6.75.

This is an attractive, well-manufactured book for the guidance of professionally trained clinicians who meet handicapped children in their practice. The greater part of the book is devoted to a description of the syndromes. An introductory section on diagnosis and a final chapter on management constitute the chief difference between this book and a standard text on mental deficiency.

I am of two conflicting opinions. First, one cannot help being irritated by an extremely careless job of editing. One can't escape the feeling that such sloppiness might carry over into the crucial material of the book. There are instances of its having done so. On the other hand, if one can ignore these faults and consider the book as a whole, and in terms of its objectives, one must say that it contains a great deal of valuable information.

The classification scheme groups all disorders under four major varieties: developmental, metabolic, neuromotor, and psychological. Each of these varieties includes several specific syndromes which are described in some detail in terms of etiology, symptoms, and, in some cases, medication or treatment. While this etio-

logic grouping would seem to be satisfactory, the reader may be confused to find deficiencies due to cerebral injury, sensory damages and nutritional disorders classified among the psychological varieties.

The author shows much greater knowledge and understanding of medical, physiologic, and pediatric matters than in psychiatric or psychological areas. The book is full of flat statements: some are misleading, some are simply wrong, and some are silly. It is hard to understand this kind of writing and editing. It will certainly alienate the professional reader who knows something of the subject and it may mislead and confuse the uninformed who most need a book of this sort.

KARL F. HEISER

*Louisville, Kentucky*

EDWARDS, ALLEN L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954. Pp. xvii+542. \$6.50.

This book is outstanding among the increasing number of texts designed to develop applied statistical competence in the perennially mathematics-free student of psychology, education, or sociology. Writing in conversational style, the author unfolds an extensive array of topics with maximum palatability and minimum sacrifice of modern statistical rationale.

The book is fairly large, containing 19 chapters followed by a 105-item bibliography, a list of 303 formulas cited in the text, an appendix of 15 tables, and answers to the excellent assortment of exercises presented in connection with each chapter. This all adds up to more than enough material for the first year of statistics.

As must be true of any introduc-

tory text, a great deal of traditional material is covered. A new book's claim to attention should be judged in terms of such criteria as clarity of exposition, self-containedness, selection of basic and specialized topics, provision of visual aids, examples, and exercises, relevance to problems of experimental design, and liaison with more advanced treatments. On all such criteria the book stands up exceptionally well. Conventional material on central tendency, variability, graphic devices, standard scores, normal curve, simple correlation and regression, other measures of association, chi square, and rudimentary analysis of variance is presented in rich detail. The only obvious omission is partial and multiple correlation and regression.

Contentwise the reviewer has only one question. This occurs in connection with Chapter 17 which deals with analysis of variance of a two-factor design. The first part of this chapter is devoted to analysis of a design involving the influence of three methods of instruction (lecture, discussion, project) upon three different kinds of achievement (a test of factual information, a test of understanding of general principles, and a test of ability to make applications). Although the tests are stated to be "comparable" it is didactically inappropriate to furnish the elementary student with an example where the dependent variable is made up of scores on three different tests. In addition the author describes the initial step in analyzing such a 3 by 3 design as testing the differences among the nine subgroups. In the example these differences are found to be significant and the analysis proceeds to the factorial form but the implication is strong that one should not go on to the factorial analysis unless the

over-all differences are found to be significant.

Over and above the usual material Edwards provides many attractive features not commonly found in the introductory text. Among these are sections on nonlinear curve fitting, the power function in tests of significance, one-tailed vs. two-tailed tests of significance, combination of tests of significance, and Tukey's procedure for comparing individual means in conjunction with analysis of variance. Perhaps the most valuable special feature is the extensive presentation of nonparametric methods, a number of which are described in the same chapter with the analogous classical method while others are discussed in a final chapter on significance tests for ranked data.

LEONARD S. KOGAN

*Institute of Welfare Research  
Community Service Society of  
New York*

LE BEAU, J. *Psycho-chirurgie et fonctions mentales*. Paris: Masson et Cie (Eds.), 1954.

This is an unusual, and in many ways, an exciting book. Written by one of France's leading brain surgeons, it deals with the general topic of psychosurgery from many angles. The author treats all the anatomical, physiological, clinical, and psychological principles involved, as well as different surgical techniques, particularly those related to selective ablations, medical complications, and postoperative treatment. Much space is given over to the results of psychosurgery in the treatment of neuroses, psychoses, mental disorders associated with epilepsy, the mental difficulties of children, and intractable pain. On all these points the author, drawing on his rich experience, has much of importance to say; his own

material is well integrated with his surveys of the literature. The main importance of this work on the surgical side derives from the author's pioneering efforts in the direction of selective surgical ablation, in which he has followed in the footsteps of Clovis Vincent.

To the psychologist, these achievements are of only peripheral interest. To him the main contribution of the book will center in those later chapters dealing with the psychological aftereffects of different types of operation. The importance of Le Beau's approach is twofold. In the first place, he does not rely, as so many other surgeons have done, on introspective reports and casual observation of the behavior of his patients. Nor does he give credence to results obtained on projective tests of low reliability and unknown validity, such as the Rorschach or the TAT. He argues firmly in favor of exact measurement in terms of objective tests measuring factorially ascertained dimensions of behavior. In this he follows the path indicated by A. Petrie in her book *Personality and the Frontal Lobes*. (Indeed, he makes specific reference to the help received from her in setting up his new psychological laboratory.) It is certainly a welcome change to find a surgeon knowledgeable enough to devote several pages to a discussion of the contributions of Thurstone, Cattell, and other factor analysts, and objective enough in his outlook to be willing to put his hypotheses to the experimental test in terms of objective personality measures.

These hypotheses themselves are of very far-reaching importance and take us back to an earlier and more hopeful stage of psychological and physiological work, when expectations were high that mental functions could be located in certain parts of the cortex. By tracing out the detailed behavioral aftereffects of operations involving different parts of the cortex, and relating these changes to personality dimensions identified in terms of factor analysis, Le Beau has given us extremely interesting hypotheses, linking such factors as extraversion, neuroticism, etc., with definite Brodmann areas (p. 380).

It would be easy to criticize the picture given by him as naive, premature, and oversimplified. All this would undoubtedly be true. Nevertheless, such feelings should not keep the reader from studying the evidence in detail. It is surprising to note to what extent many apparently unrelated factors fall into place when seen in terms of the scheme presented by Le Beau. In any case, whatever the faults of the scheme, it possesses the outstanding advantage of being definite enough to permit of disproof; many deductions can be made from it which are experimentally testable. At an early stage of the development of a theory it would be unfair to ask for more than this: that it should unify known facts and predict unknown facts. It is to be hoped that this book will be widely read, and, equally important, that it be translated into English.

H. J. EYSENCK

University of London





### IMPORTANT ANNOUNCEMENT

As approved by the Board of Directors and the Council of the APA, beginning with January 1956, a new journal will be published by the Association. This journal will review books, monographs, films, and related publications—a function presently performed by four different APA journals.

Accordingly, book reviews will appear in the *Psychological Bulletin* only through the completion of the present volume, 82, November 1955 issue.

Hereafter all publications submitted for review and requests to prepare reviews should be directed to the editor of the new journal:

*Contemporary Psychology, A Journal of Reviews*

E. G. Boring, Editor

Memorial Hall

Harvard University

Cambridge 38, Massachusetts

The journal will appear monthly and the subscription price (for non-APA members) will be \$8.00. Subscription orders should be sent to:

American Psychological Association

1333 Sixteenth Street N.W.

Washington 6, D.C.